Scientific and Educational Eastern European Medical Cluster

MEDICAL ETHICS

SCIENTIFIC INTERDISCIPLINARY JOURNAL OF YAROSLAVL STATE MEDICAL UNIVERSITY
AND PIROGOV RUSSIAN NATIONAL RESEARCH MEDICAL UNIVERSITY

SUBMISSION https://medet.rsmu.press/
CORRESPONDENCE https://medet.rsmu.press/
COLLABORATION editor@rsmu.press
ADDRESS Ostrovityanov Street 1, Moscow, 119997, Russia

Indexed in RSCI

Open access to archive

Issue DOI: 10.24075/medet.2025-02
The mass media registration certificate серия ПИ № ФС77-81021 от 02 июня 2021 г.
Founders: Yaroslavl State Medical University (Yaroslavl, Russia)
Pirogov Russian National Research Medical University (Moscow, Russia).
Publisher: Pirogov Russian National Research Medical University; adress: Ostrovityanov Street 1, Moscow, 117997, Russia

Approved for print 30.06.2025
Circulation: 100 copies. Printed by Print.Formula
www.print-formula.ru

Научно-образовательный медицинский кластер Восточно-Европейский

МЕДИЦИНСКАЯ ЭТИКА

НАУЧНЫЙ МЕЖДИСЦИПЛИНАРНЫЙ ЖУРНАЛ
ЯРОСЛАВСКОГО ГОСУДАРСТВЕННОГО МЕДИЦИНСКОГО УНИВЕРСИТЕТА
И РОССИЙСКОГО НАЦИОНАЛЬНОГО ИССЛЕДОВАТЕЛЬСКОГО МЕДИЦИНСКОГО
УНИВЕРСИТЕТА ИМ. Н. И. ПИРОГОВА

ПОДАЧА РУКОПИСЕЙ https://medet.rsmu.press/
ПЕРЕПИСКА С РЕДАКЦИЕЙ https://medet.rsmu.press/
СОТРУДНИЧЕСТВО editor@rsmu.press
АДРЕС РЕДАКЦИИ ул. Островитянова, д.1, г. Москва, 119997

Журнал включен в РИНЦ

Здесь находится открытый архив журнала

DOI выпуска: 10.24075/medet.2025-02
Свидетельство о регистрации средства массовой информации серия ПИ № ФС77-81021 от 02 июня 2021 г.
Учредители: ФГБОУ ВО «Ярославский государственный медицинский университет» Минздрава России (Ярославль, Россия);
ФГАОУ ВО «Российский национальный исследовательский медицинский университет имени Н.И. Пирогова» Минздрава России (Москва, Россия).
Издатель: ФГАОУ ВО «Российский национальный исследовательский медицинский университет имени Н.И. Пирогова» Минздрава России;
адрес: ул. Островитянова, д.1, г. Москва, 117997, Россия, 8(495)434-03-29

Подписано в печать 30.06.2025
Тираж 100 экз. Отпечатано в типографии Print.Formula
www.print-formula.ru

MEDICAL ETHICS **2, 2025**
МЕДИЦИНСКАЯ ЭТИКА

Contents
Содержание

BIOETHICAL PROBLEMS OF QUANTUM COMPUTING

Khokhlov AL[1], Pavlov AV[1], Kotlovsky MYu[1,2] ✉, Potapov MP[1], Gabidullina LF[1]

[1] Yaroslavl State Medical University, Yaroslavl, Russia

[2] National Research Institute of Public Health named after NA Semashko, Moscow, Russia

Today, quantum computing is one of the most promising areas of modern science that can solve the problems which are too hard to handle for classical computers. However, a huge potential accompanied by a rapid progress in this area creates significant ethical risks, including threat of hacking existing cryptographic systems and the problem of non-transparent decisions made using quantum computing. The article highlights the need for timely ethical analysis, development of a regulatory framework, and interdisciplinary collaboration to ensure sustainable and socially acceptable technology development. Main ethical, bioethical challenges, in particular, that arise against the background of quantum computing, and identification of possible ways to regulate them captivate special attention.

Keywords: quantum computing, bioethics

Author contribution: A. L. Khokhlov — formulation of the problem, discussion of key ethical issues, planning and discussion of the article; A. V. Pavlov — study of the literature on the topic, systematization and generalization of data, participation in the discussion of the results, writing and design of the article; M. Yu. Kotlovsky, M. P. Potapov, L. F. Gabidullina — study of the literature on the topic, participation in the discussion of the results and writing of the article.

✉ **Correspondence should be addressed**: Mikhail Yu. Kotlovsky
Revolutsionnaya str., 5, Yaroslavl, 150000, Russia; m.u.kotlovskiy@mail.ru

Received: 05.05.2025 **Accepted:** 19.05.2025 **Published online:** 23.06.2025

DOI: 10.24075/medet.2025.007

БИОЭТИЧЕСКИЕ ПРОБЛЕМЫ КВАНТОВЫХ ВЫЧИСЛЕНИЙ

А. Л. Хохлов[1], А. В. Павлов[1], М. Ю. Котловский[1,2] ✉, М. П. Потапов[1], Л. Ф. Габидуллина[1]

[1] Ярославский государственный медицинский университет, Ярославль, Россия

[2] Национальный НИИ общественного здоровья имени Н. А. Семашко, Москва, Россия

На сегодняшний день квантовые вычисления представляют собой одно из наиболее перспективных направлений развития современной науки, способное решать задачи, не доступные классическим системам. Однако, наряду с огромным потенциалом, стремительный прогресс в этой области порождает существенные этические риски, включая угрозу взлома существующих криптографических систем и проблему непрозрачности принимаемых с помощью квантовых вычислений решений. В статье подчеркивается необходимость своевременного этического анализа, разработки нормативно-правовой базы и междисциплинарного сотрудничества для обеспечения устойчивого и общественно приемлемого развития технологии. Особое внимание уделяется рассмотрению основных этических, в частности биоэтических, вызовов, возникающих на фоне развития квантовых вычислений, и определению возможных путей их регулирования.

Ключевые слова: квантовые вычисления, биоэтика

Вклад авторов: А. Л. Хохлов — постановка проблемы, обсуждение ключевых этических вопросов, планирование и обсуждение статьи; А. В. Павлов — изучение литературы по теме, систематизация и обобщение данных, участие в обсуждении результатов, написание и оформление статьи; М. Ю. Котловский, М. П. Потапов, Л. Ф. Габидуллина — изучение литературы по теме, участие в обсуждении результатов и написание статьи.

✉ **Для корреспонденции**: Михаил Юрьевич Котловский
ул. Революционная, д. 5, г. Ярославль, 150000, Россия; m.u.kotlovskiy@mail.ru

Статья поступила: 05.05.2025 **Статья принята к печати:** 19.05.2025 **Опубликована онлайн:** 23.06.2025

DOI: 10.24075/medet.2025.007

INTRODUCTION

Today, quantum computing is one of the most promising areas of modern science. They are developed using fundamental quantum properties of matter such as superpositions, entanglements and interference for information processing. Owing to that, quantum computers are able to perform tasks that are either extremely difficult or completely impossible for classical systems of today. This applies, in particular, to complex chemical process modeling, accelerated search in big data, as well as cryptographic analysis.

The current stage in quantum computing development represents a transition to a new technological order, especially in cryptography, artificial intelligence, and biomedical research. Potential advantages of the new approach are also accompanied by significant risks, many of which are of ethical origin. The history of information technology shows that ignoring moral and social consequences of scientific discoveries can lead to massive issues such as privacy violations, spread of misinformation, or growth of digital inequality in society.

Currently, possible use of quantum algorithms to crack existing cryptographic information security systems is one of the most significant threats. We are also concerned about transparency of decisions made using quantum computing. This is especially important in areas related to human life and health, such as medicine or forensic science. All this greatly increases the need for timely and adequate ethical analysis.

Ethical issues related to quantum computing cover a variety of aspects, from principles of equity and accountability to implications for global security. To solve the issues, a clear regulatory framework and ethical guidelines that can ensure sustainable and socially acceptable technology development are required.

Thus, ethics of quantum computing is becoming an integral part of scientific and engineering thinking of the 21st century. Cooperation between specialists from various fields is necessary for its development. The specialists are physicists, computer scientists, lawyers, and philosophers. Active involvement of civil society institutions is required. In this chapter, it is proposed to consider the main bioethical challenges that arise during rapid progress of this field of science, and identify possible ways to regulate them.

1. FUNDAMENTALS OF QUANTUM COMPUTING: KEY DIFFERENCES AND OPPORTUNITIES

Quantum computing is a special approach to information processing based on the principles of quantum mechanics [1]. Unlike standard classical machines that use bits, which can represent values of 0 or 1, quantum computers operate with qubits, which can exist in multiple states at once due to superposition. It provides quantum systems with a unique opportunity to perform parallel calculations. Traditional processors cannot do that.

1.1. The main differences of quantum calculations

Origin of data representation. Classical computers use bits to encode information. Quantum systems use qubits, which can take not only the values 0 or 1, but also their linear combinations. This is how their computing capabilities can be numerously increased.

Parallelism. Quantum systems can simultaneously analyze numerous options owing to superimposition. Due to that, certain types of problems such as iteration or optimization can be solved faster.

Entanglement. Quantum entanglement allows you to connect qubits in such a way that changing one instantly leads to a change in the other. This powerful tool allows to build complex algorithms.

Algorithmic advantage. Shor's and Grover's algorithms show that quantum computers can significantly outperform classical ones [2]. They are especially efficient in factorization, search, and modeling of chemical processes.

Physical instability. Qubits are extremely sensitive to external influences and lead to decoherence. It is difficult to maintain quantum states as ultra-low temperatures and error correction systems are required. This makes quantum computing technically complex and cost-effective.

1.2. Types of quantum devices

At the present stage of development, quantum devices are mainly represented by quantum computers and quantum simulators:

- universal quantum computers can perform a wide range of tasks and simulate any quantum system. They have a flexible architecture and implement a universal set of quantum logic operations;
- quantum simulators are often tailored to model specific physical systems such as modeling molecular interactions. In these areas, they are often more efficient than universal machines.

1.3. Potential and limitations

Quantum computing is highly potential in solving the problems that cannot be solved by classical systems, such as modeling molecular interactions, cryptographic analysis, optimization, and machine learning. Quantum parallelism, entanglement, and the possibility of exponential acceleration when executing certain algorithms are key advantages of quantum computers. However, implementation of this potential is limited by a number of factors.

First, significant difficulties can be seen at the level of hardware: qubits are extremely unstable, subject to decoherence, and operate at ultra-low temperatures and in the presence of complex technical infrastructure. Existing quantum devices are still laboratory prototypes with limitations in qubit count and error rates, which makes it necessary to use complex correction systems.

Second, there are significant limitations associated with development of quantum algorithms. To date, there exist only a limited number of algorithms with proven quantum superiority (for example, Shor's and Grover's algorithms), and their use is limited to a narrow class of tasks. Designing new algorithms requires a strong foundation in quantum physics and mathematics, as well as mastering new programming languages and computational models that are under active development [3]. Meanwhile, most of the known algorithms show practical efficiency primarily with a large number of qubits, which cannot be achieved with the current level of technology. It is expected that they will appear in future.

Since universal quantum computers are still far from practical application, **hybrid quantum-classical approaches** (for example, Variational Quantum Algorithms) have the greatest prospects. However, their implementation requires deep integration of two types of systems and novel engineering solutions [4].

Thus, despite their outstanding potential, quantum technologies are still at the stage of scientific and engineering testing, and their practical application requires overcoming both technical and methodological barriers.

2. POTENTIAL THREATS AND RISKS

2.1. Digital security threats

Quantum technologies can solve problems which cannot be overcome by classical processors. This, however, poses serious threats to existing data encryption systems. Modern cryptographic methods (for example, RSA) are based on complex big number factorization, but the advent of scalable quantum computers using Shor's algorithm makes the task solvable in polynomial time [5].

The strategy of individuals trying to gain unauthorized access to 'store now, decrypt later' information is of particular concern. Meanwhile, the encrypted data is intercepted and deposited for subsequent decryption, at a time when quantum computing reaches the required power. This is especially critical for information with long-term confidentiality, such as documents containing information classified as a state secret, medical archives, and personal correspondence.

In this regard, transition to post-quantum cryptography based on principles resistant to quantum attacks is required to protect digital infrastructure. The solutions can be effective only with international cooperation and timely revision of cryptographic standards.

2.2. Military and political risks

Quantum technologies can reshape global power dynamics significantly. Countries with access to quantum computing will gain significant advantages in intelligence, data analysis, information security, and defense system development. The capabilities of these future systems will involve modeling tactical

scenarios, optimization of combat operations, and developing cyber-warfare capabilities.

Similar to nuclear development, the quantum arms race can significantly heighten international instability. Possible militarization of technology will be accompanied by growth of closed projects, reducing their transparency and hampering control. Concentration of all quantum powers in a limited number of countries can increase inequality and disturb the balance in the global security system.

The risks can be struggled by implementing international agreements aimed at prevention of uncontrolled proliferation and use of quantum systems for purely militaristic purposes.

2.3. Risk of social inequality

Development of quantum computing can exacerbate existing digital and economic inequalities [6]. Access to advanced technologies is most often limited by large scientific centers, corporations, and economically developed superpower states. This threatens to widen the global gap between the "technological center" and the developing periphery.

To exclude the threat, an inclusive scientific policy is necessary. It means expansion of educational programs, provision of open access to research results, and support of infrastructure in countries with limited resources. Patent regulation is also expected to be important, as excessive monopolization of quantum developments may hinder equal access to new knowledge and achievements.

Functioning patent systems should facilitate distribution of useful technologies without creating artificial barriers. Meanwhile, transparency in developments and publications is also critical. It will reduce the risks of manipulation, increase trust, and turn quantum technologies into the subject of public control.

Modern international norms, including agreements on intellectual property rights and open data, need to be adapted to new challenges of the quantum era. Ethical and legal support is necessary to create a fair technological environment where innovation serves the interests of all countries and communities around the world.

3. RESPONSIBLE DEVELOPMENT AND USE

Development of quantum computing requires ethical and legal support to ensure a balance between scientific progress and interests of the society. For ethical development of quantum technologies, it is necessary to rely on the principles of precaution, transparency, accessibility and open cooperation. These provisions will serve as an ethical guideline and allow integration of quantum technologies into public and scientific processes without loss of trust and fairness.

The precautionary principle. If the technology poses potential threats, it should be used along with a preliminary assessment of all possible risks during its implementation. In terms of quantum computing, this is especially important in the fields of security, international relations, and civil rights. Prior to mass implementation of such solutions, it is necessary to perform a comprehensive ethical and legal examination.

Transparency. Effective and responsible use of quantum technologies will require an open and understandable description of algorithms, development goals, and possible consequences of using these technologies. Transparency is especially important when quantum solutions are used in critical areas such as medicine, defense, and digital security. Availability of information will help increase trust and reduce the risks of abuse from individual players.

Inclusivity. When technologies are developed, the interests of a wide range of participants, states, research institutes, small research teams, and developing countries should be taken into account. Access to resources, education, knowledge, and research results should not be limited only to economically developed regions or multinational corporations.

Openness. Access to scientific publications, source codes, and research protocols is an important condition for progress in ethics. Modern open platforms such as Qiskit (IBM) or PennyLane (Xanadu) demonstrate that collaboration and knowledge sharing are possible. This will reduce barriers of entry into this scientific field and contribute to a more even development of quantum technologies.

4. THE ROLE OF ORGANIZATIONS IN SHAPING THE ETHICAL AGENDA

Universities and research centers. Higher educational institutions all over the world are not only developing technologies, but also shaping the worldview of future specialists working with them. Educational programs in ethics of technology, interdisciplinary research and introduction of standards of scientific integrity are the most important elements of sustainable technological development of society. Besides, Universities are actively initiating research on the ethical aspects of artificial intelligence, quantum computing, and other breakthrough fields.

The private sector. Companies, especially leading technology corporations, are increasingly integrating ethical standards into their daily operations through corporate and social responsibility (CSR) mechanisms. It means establishment of internal ethics committees, publication of codes of ethics, and participation in international initiatives aimed at ethical regulation. The leading market players also play a key role in standardization of new technologies becoming drivers of ethics implementation in practice.

International initiatives and a consortium. To ensure safe development of quantum computing, these structures form the basis for an international dialogue and a regulatory framework. They monitor compliance with ethical principles, develop accountability mechanisms, and assess the risks of exploiting these technologies.

- **Institute of Electrical and Electronics Engineers (IEEE)**: an international organization that develops technical and ethical standards. The *IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems* also addresses quantum computing.
- **Quantum Ethics Consortium**: an interdisciplinary platform for developing ethical approaches to application of quantum technologies. It ensures cooperation of scientists, engineers, and policy makers in this area.
- **AI Ethics Guidelines Global Initiative**: an international movement whose activities are aimed at developing universal standards for breakthrough technologies. It includes participation of the United Nations, the WIPO (World Intellectual Property Organization) and other private organizations [7].

Thanks to cooperation between universities, businesses and interstate structures, it is possible to form global standards that are selectively adapted to particular regional peculiarities and needs.

5. THE NEED FOR REGULATION AND INTERNATIONAL COOPERATION

Quantum computing has a high modernization potential that can affect key areas of public life, from national security and economics to medicine and sensitive data management.

This progress affects not only scientific, but also geopolitical interests, and requires well-thought-out regulatory mechanisms, global interaction, and creation of an adequate system of checks and balances.

The need for international regulation of quantum technology development in this regard is due to:

- threat of destroying the existing global cryptographic infrastructure and undermining cybersecurity;
- risk of monopolization of technology by individual states (superpowers) or multinational corporations;
- increasing the likelihood of technological inequality between countries;
- possibility of militarization and entering a new round of the arms race using the established foundation of quantum developments;
- lack of developed ethical guidelines and global norms necessary for protection of human rights and individual freedom.

In this case, **geopolitical competition** and desire for technological leadership can act as a barrier to global regulation, which is reflected in **secrecy of defense developments**, which hinders transparency and control. This is driven by **uneven development** of quantum technologies in different countries around the globe [8]. **Differences in legal systems** and limited sovereignty of individual states are essential as well.

Activity of the international organizations below is given as an example of international initiatives in this field:

- **WEF (World Economic Forum)** builds a "quantum economy" with an emphasis on sustainability and global ethics;
- **UNESCO** is discussing formation of an international code of ethics for quantum technologies based on rich bioethical developments;
- **OECD** forms recommendations on responsible implementation of breakthrough technologies and stimulates intersectoral dialogue;
- **Quantum Flagship (EU)** supports standardization, ethical control and cross-border cooperation within the European research policy [9].
- It seems the above barriers can be cleared with the following methods:
- **soft law** [10] — creation of codes, declarations and framework agreements that form a culture of responsibility between the leading players working in this field;
- **scientific diplomacy** — development of international research consortia and open networks;
- **inclusivity in technology** — ensuring equal access to education and infrastructure for people from all countries;
- **open standards platform** — use of publicly available quantum instruments under international jurisdiction [11];
- **interdisciplinary approach** — involving lawyers, philosophers, engineers and representatives of civil society in discussion of ethical issues of quantum technology development.

Thus, international cooperation in quantum technologies is not an optional, but a vital condition for preventing future global conflicts, building trust and fairly distributing the results of scientific progress within society. Development of ethically sound global regulatory mechanisms should go hand in hand with scientific advances and include both initiatives from states, international institutions, research centers, and individual citizens.

CONCLUSION

Quantum computing promises a qualitative breakthrough in science, medicine, industry, communications and many sectors of national economy. This opens up new horizons for solving problems that have long been considered inaccessible to classical computing systems. Meanwhile, such a large-scale human progress is accompanied by serious challenges to society [12]. These threats range from information security to the risks of increasing global strategic inequality and, as a result, global military and political instability.

Ethical aspects of quantum technology development and application hold a specific place in modern scientific discourse [13]. It requires a systematic approach that unites the efforts of physicists, engineers, lawyers, philosophers and politicians. Meanwhile, the principles of precaution, transparency, inclusivity and openness should form the basis for ethically sustainable development in this field.

Formation of an ethical culture among specialists working with quantum technologies is becoming one of the key tasks of modern scientific training. Ethical literacy, as part of professional training, is necessary to ensure a balance between technological progress and public interests.

The future of quantum technologies should be based on the principles of international partnership, trust and equal access to these technologies. Safe, fair and benefit-oriented development of such systems is possible only with joint efforts.

References

1. Kitaev AYu, Shen` AN, Vyaly`j MN. Klassicheskie i kvantovy`e vy`chisleniya. M. MCzNMO. 1999; 320 s. Russian.
2. Valiev KA. Kvantovy`e komp`yutery`: nadezhdy` i real`nost`. M. Nauka. 2001; 256 s. Russian.
3. Antipenko LG. Ponyatie kvantovoj logiki: matematicheskie i filosofskie aspekty`. Vestnik Moskovskogo universiteta. Seriya 7: Filosofiya. 2015; (3): 3–15. Russian.
4. Nil`sen MA, Chuang IL. Kvantovy`e vy`chisleniya i kvantovaya informaciya. M. Vil`yams. 2020; 800 s. Russian.
5. Aaronson S. Kvantovy`e vy`chisleniya so vremyon Demokrita. Per. s angl. N. Lisovoj. Nauch. red. A. L`vovskij. M. Al`pina non-fikshn. 2018; 494 s. Russian.
6. Kaku M. Kvantovoe prevosxodstvo: revolyuciya v vy`chisleniyax, kotoraya izmenit vsyo: per. s angl. M. Al`pina non-fikshn. 2021; 352 s. Russian.
7. Borzunov SV, Kurgalin SD. Osnovy` kvantovy`x vy`chislenij i kvantovoj teorii informacii: uchebnoe posobie. SPb. BXV-Peterburg. 2022; 144 s. Russian.
8. Xidari Dzh. Kvantovy`e vy`chisleniya. Prikladnoj podxod: per. s angl. M. DMK Press. 2021; 352 s. Russian.
9. Perrier E. Ethical Quantum Computing: A Roadmap. arXiv preprint arXiv: 2102.00759. 2021.
10. Schumacher B, Westmoreland MD. Quantum Processes, Systems, and Information. Cambridge: Cambridge University Press. 2010; 500 p.
11. Wilde MM. Quantum Information Theory. Cambridge: Cambridge University Press. 2013; 685 p.
12. EITC/QI/QIF Osnovy` kvantovoj informacii. Akademiya EITCA. 2021.
13. Kvantovaya i prikladnaya e`tika. SystemsWorld Club. 2022. Available from URL: https://systemsworld.club/t/topic/3029?sysclid=ma2tqu2wx1885516128 (accessed: 15.04.2025). Russian.

Литература

1. Китаев А. Ю., Шень А. Н., Вялый М. Н. Классические и квантовые вычисления. М. МЦНМО. 1999; 320 с.
2. Валиев К. А. Квантовые компьютеры: надежды и реальность. М. Наука. 2001; 256 с.
3. Антипенко Л. Г. Понятие квантовой логики: математические и философские аспекты. Вестник Московского университета. Серия 7: Философия. 2015; (3): 3–15.
4. Нильсен М. А., Чуанг И. Л. Квантовые вычисления и квантовая информация. М. Вильямс. 2020; 800 с.
5. Ааронсон С. Квантовые вычисления со времен Демокрита. Пер. с англ. Н. Лисовой. Науч. ред. А. Львовский. М. Альпина нон-фикшн. 2018; 494 с.
6. Каку М. Квантовое превосходство: революция в вычислениях, которая изменит все: пер. с англ. М. Альпина нон-фикшн. 2021; 352 с.
7. Борзунов С. В., Кургалин С. Д. Основы квантовых вычислений и квантовой теории информации: учебное пособие. СПб. БХВ-Петербург. 2022; 144 с.
8. Хидари Дж. Квантовые вычисления. Прикладной подход: пер. с англ. М. ДМК Пресс. 2021; 352 с.
9. Perrier E. Ethical Quantum Computing: A Roadmap. arXiv preprint arXiv: 2102.00759. 2021.
10. Schumacher B, Westmoreland MD. Quantum Processes, Systems, and Information. Cambridge: Cambridge University Press. 2010; 500 p.
11. Wilde MM. Quantum Information Theory. Cambridge: Cambridge University Press. 2013; 685 p.
12. EITC/QI/QIF Основы квантовой информации. Академия EITCA. 2021.
13. Квантовая и прикладная этика. SystemsWorld Club. 2022. Режим доступа: [Электронный ресурс] URL: https://systemsworld.club/t/topic/3029?ysclid=ma2tqu2wx1885516128 (дата обращения: 15.04.2025).

THE CONCEPT OF PREDICTABLE HARM IN DEVELOPMENT OF AI-POWERED MEDICAL DEVICES

Begishev IR ✉, Shutova AA

Kazan Innovative University named after VG Timiryasov, Kazan, Russia

The article reviews the concept of predictable harm as a methodological tool for a comprehensive risk assessment while developing and implementing AI-powered medical devices. The study is relevant due to exponential growth of using AI-powered technologies in healthcare and lack of unified approaches to prediction of potential negative consequences of their usage. Existing regulatory approaches to risk assessment, including Russian regulatory documents and international standards, have been analyzed. A multidimensional classification of types of predictable harm is proposed considering the entire life cycle of medical AI systems. Special attention is given to ethical aspects of using artificial intelligence in medicine, including the principles of patient autonomy, equity, non-harm and transparency of algorithms. An expanded matrix for assessing predictable harm has been developed. It integrated technological, clinical and ethical parameters for each stage of development and implementation of AI systems in medical practice. The results of the study can be used as a methodological framework for developers of medical AI systems, regulatory authorities and medical organizations in assessing safety and effectiveness of introducing intelligent technologies into clinical practice.

Keywords: artificial intelligence, medical devices, predictable harm, ethics of artificial intelligence, regulation, patient safety, risk management

Financing: the work was carried out under the grant from the Academy of Sciences of the Republic of Tatarstan, provided to young candidates of sciences (postdoctoral fellows) so that they could defend their doctoral dissertation, perform research and labor functions in scientific and educational organizations of the Republic of Tatarstan within the framework of the Scientific and Technological Development of the Republic of Tatarstan State Program.

Acknowledgements: the authors express their sincere gratitude to the Academy of Sciences of the Republic of Tatarstan for financial support of this study based on the results of a competitive selection (grant No. 153/2024-PD dated December 16, 2024) for 'Ensuring the technological sovereignty of the healthcare system by criminal means' research project, as well as deep gratitude to the staff of the Scientific Research Institute of Digital Technologies and Law of Kazan Innovative University named after Timiryasov VG for valuable consultations and constructive discussion of the conceptual provisions of the study.

Author contribution: Begishev IR — research conceptualization; development of predictable harm theoretical foundations; analysis of specific risks associated with using artificial intelligence in medical devices; systematization of predictable harm typology; research of legal framework for regulating AI systems in healthcare with various jurisdictions; formulation of research conclusions; preparation of the manuscript initial version; Shutova AA — development of a methodology for predicting and preventing predictable harm; creation of a matrix for assessing predictable harm at various stages of medical AI system life cycle; developing recommendations for practical implementation of predictable harm concept; analysis of literature sources; editing and critical revision of the manuscript with introduction of valuable intellectual content; visualization of research (making tables).

Compliance with ethical standards: a meeting of the ethics committee was not required, since this study is theoretical and methodological in nature and analyzes open literature sources and regulatory legal documents, without conducting experiments involving humans or animals and without using personal patient data.

✉ **Correspondence should be addressed:** Ildar R. Begishev
42 Moskovskaya St., Kazan, Russia, 420111; begishev@mail.ru

Received: 03.05.2025**Accepted:** 16.05.2025 **Published online:** 20.06.2025

DOI: 10.24075/medet.2025.006

КОНЦЕПЦИЯ «ПРЕДСКАЗУЕМОГО ВРЕДА» ПРИ РАЗРАБОТКЕ МЕДИЦИНСКИХ ИЗДЕЛИЙ НА ОСНОВЕ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА

И. Р. Бегишев ✉, А. А. Шутова

Казанский инновационный университет имени В. Г. Тимирясова, Казань, Россия

В статье представлена концепция «предсказуемого вреда» как методологический инструмент для комплексной оценки рисков при разработке и внедрении медицинских изделий на основе искусственного интеллекта. Актуальность исследования обусловлена экспоненциальным ростом применения ИИ-технологий в здравоохранении при одновременном отсутствии унифицированных подходов к прогнозированию потенциальных негативных последствий их использования. Проведен критический анализ существующих регуляторных подходов к оценке рисков, включая отечественные нормативные документы и международные стандарты. Предложена многомерная классификация типов предсказуемого вреда с учетом всего жизненного цикла медицинских ИИ-систем. Особое внимание уделено этическим аспектам применения искусственного интеллекта в медицине, включая принципы автономии пациента, справедливости, непричинения вреда и прозрачности алгоритмов. Разработана расширенная матрица оценки предсказуемого вреда, интегрирующая технологические, клинические и этические параметры для каждого этапа разработки и внедрения ИИ-систем в медицинскую практику. Результаты исследования могут служить методологической основой для разработчиков медицинских ИИ-систем, регуляторных органов и медицинских организаций при оценке безопасности и эффективности внедрения интеллектуальных технологий в клиническую практику.

Ключевые слова: искусственный интеллект, медицинские изделия, предсказуемый вред, этика искусственного интеллекта, регулирование, безопасность пациентов, управление рисками

Финансирование: работа выполнена за счет гранта Академии наук Республики Татарстан, предоставленного молодым кандидатам наук (постдокторантам) с целью защиты докторской диссертации, выполнения научно-исследовательских работ, а также выполнения трудовых функций в научных и образовательных организациях Республики Татарстан в рамках Государственной программы Республики Татарстан «Научно-технологическое развитие Республики Татарстан».

Благодарности: авторы выражают искреннюю признательность Академии наук Республики Татарстан за оказанную финансовую поддержку данного исследования по результатам конкурсного отбора (грант № 153/2024-ПД от 16 декабря 2024 г.) для научно-исследовательского проекта «Обеспечение технологического суверенитета системы здравоохранения уголовно-правовыми средствами», а также глубокую благодарность коллективу Научно-исследовательского института цифровых технологий и права Казанского инновационного университета имени В. Г. Тимирясова за ценные консультации и конструктивное обсуждение концептуальных положений исследования.

Вклад авторов: И. Р. Бегишев — концептуализация исследования; разработка теоретических основ понятия «предсказуемого вреда»; анализ специфики рисков, связанных с применением искусственного интеллекта в медицинских изделиях; систематизация типологии предсказуемого вреда; исследование

нормативных правовых основ регулирования ИИ-систем в здравоохранении в различных юрисдикциях; формулирование выводов исследования; подготовка первоначального варианта рукописи; А. А. Шутова — разработка методологии прогнозирования и превенции предсказуемого вреда; создание матрицы оценки предсказуемого вреда на различных этапах жизненного цикла медицинских ИИ-систем; формирование рекомендаций по практической имплементации концепции предсказуемого вреда; анализ источников литературы; редактирование и критический пересмотр рукописи с внесением ценного интеллектуального содержания; визуализация исследования (разработка таблиц).

Соблюдение этических стандартов: заседание этического комитета не требовалось, поскольку данное исследование носит теоретико-методологический характер и основано на анализе открытых литературных источников и нормативных правовых документов, без проведения экспериментов с участием людей или животных и без использования персональных данных пациентов.

✉ **Для корреспонденции:** Ильдар Рустамович Бегишев
ул. Московская, д. 42, Республика Татарстан, г. Казань, 420111; begishev@mail.ru

Статья поступила: 03.05.2025 **Статья принята к печати:** 16.05.2025 **Опубликована онлайн:** 20.06.2025

DOI: 10.24075/medet.2025.006

Integration of artificial intelligence into medical devices offers unprecedented opportunities to improve diagnostic processes, personalize therapeutic approaches, and optimize clinical solutions. However, rapid introduction of AI systems into healthcare is associated with specific risks that require systematic analysis and proactive management. In this context, the concept of predictable harm is gaining crucial significance as a methodological tool for preventive identification and minimization of potential negative consequences of using AI-powered medical devices. A critical analysis of existing regulatory approaches to assessing the risks of AI systems in healthcare, as well as integration of ethical principles into the process of forecasting and preventing possible harm is of particular importance.

The relevance of the study is determined by exponential growth of the market for AI solutions in healthcare and lack of unified approaches assessing their safety. According to the Grand View Research analytical report, the global market for artificial intelligence in medicine will reach 120.2 billion US dollars by 2028 with an annual increase of about 41.8%. It shows the scope of challenges in the field of patient safety [1].

The purpose of this study is to form a methodological framework for identifying and minimizing predictable harm when developing and implementing AI-powered medical devices.

To achieve this goal, the following tasks have been set:

1. To conceptualize the term of predictable harm in the context of medical AI technologies;
2. To analyze the specifics of the risks associated with the use of artificial intelligence in medical devices;
3. To investigate existing approaches to regulation of safety of AI systems in healthcare and compare them with the author's concept;
4. To develop a methodology for predicting and preventing potential harm when creating medical AI systems with detailed ethical elaboration.

ESSENTIAL PART

The concept of predictable harm for AI-powered medical devices is a methodological construct that integrates the principles of predictive risk analysis, proactive safety management, and iterative reassessment of the potential harmful effects of the technology. The fundamental difference of this concept from traditional approaches to risk assessment is in the shift of focus from reactive incident response to preventive forecasting of possible scenarios of adverse events caused by specific functioning of AI systems.

In the reviewed context, the terminological definition of predictable harm can be formulated as a set of potential negative effects of using medical AI systems. It is possible to identify and minimize them systematically analyzing characteristics of the technology, the context of its application and possible trajectories of evolution of the system during operation. The key attributes of this definition include predictive nature of assessment, a systematic approach to risk analysis, and consideration of the dynamic nature of AI technologies.

At the present stage, there are some regulatory approaches to risk assessment in the use of AI-powered medical devices.

Thus, Order No. 686n of the Ministry of Health of the Russian Federation dated July 7, 2020 [2] and letter No. 02I-297/20 of the Federal Service for Healthcare Supervision dated February 13, 2020 [3] provide for risk rating. According to it, all AI-powered medical devices are classified as Class III MD before their application and at the stage of state registration. This approach is aimed at centralized regulation and a priori high-risk classification of all AI systems in medicine.

The International Forum of Medical Device Regulators (IMDRF, 2014) offers a differentiated classification of potential risks of AI-powered medical devices, depending on clinical application and possible impact on the treatment process (Software as a Medical Device: Possible Framework for Risk Categorization and Corresponding Considerations) [4]. This classification takes into account both the severity of potential harm to the patient and the role the AI system plays in the clinical process.

The industry appendix to the Code of Ethics for Artificial Intelligence of Alliance for Artificial Intelligence provides details for gradation of risks depending on the severity of errors associated with the use of artificial intelligence systems and focuses on consequences of incorrect medical decisions [5].

Specific risks associated with the use of artificial intelligence in medical devices are due to a number of unique characteristics of these technologies: autonomous functioning, potential non-transparency of the decision-making process (the black box problem), capacity to self-education and adaptation, and high dependence on source data quality [6]. These features constitute a multidimensional risk profile that needs a differentiated approach to their identification and management.

The proposed concept of predictable harm is characterized by a multidimensional structure focused on the entire life-cycle of AI systems. It includes as follows:

– proactive focus on risk identification;

Table 1. Typology of predictable harm for medical AI systems

Harm category	Distinguishing feature	Examples	Identification and minimization methods
Algorithmic	It is associated with defects in mathematical models and decision-making logic	False positive/ false negative results, classification errors	Validation of representative samples, testing of boundary cases
Data-centric	Caused by problems in the training data	Systematic bias, inapplicability to certain groups of patients	Audit of data, stratification of samples, control of representativeness
Integrative	It occurs when the AI system interacts with clinical processes	Violations of protocols, conflicts with existing systems	Process modeling, simulation of clinical scenarios
Interpretative	Associated with incorrect interpretation of results by users	Overestimation/underestimation of AI recommendations, omission of critical information	Interface optimization, user training
Evolutionary	It is driven by the change in the system behavior during self-education	Concept drift, accuracy degradation	Performance monitoring, periodic recertification

- differentiated approach to types of harm (algorithmic, data-centric, integration, etc.);
- multilevel stratification of responsibilities of participants;
- iterative risk assessment and adaptation to the evolution of AI systems;
- integration of technological and clinical aspects of quality.

Table 1 systematizes the types of predictable harm for medical AI systems.

Safety regulation of AI-powered medical devices is characterized by significant heterogeneity of approaches in different jurisdictions. The European Union is implementing a structured regulatory system through the Medical Devices Regulation (MDR 2017/745) [7] and the Artificial Intelligence Regulation (AI Act) [8], which classifies medical AI systems as high-risk ones and sets strict requirements for their transparency and validation.

The US Food and Drug Administration (FDA) implements an adaptive approach which is based on the Pre-Certification Program focusing on evaluation of development processes and quality culture of the developer [9]. This approach involves continuous monitoring of system performance under real operating conditions and iterative reassessment of the risk-benefit profile.

Based on the analysis of existing approaches and regulatory requirements, an integrated methodology for predicting and preventing predictable harm when developing AI-powered medical devices is proposed, which includes the following components: a multi-level risk assessment model, an inclusive validation system for heterogeneous populations of patients, mechanisms for ensuring interpretability of algorithms, an infrastructure for continuous performance monitoring, and iterative safety reassessment processes.

The proposed methodology can be implemented in practice through the matrix of predictable harm assessment presented in Table 2.

Ethical aspects form an integral part of the predictable harm concept and are reflected at all stages of AI-powered medical device life cycle. Let's look at the main dimensions of ethical responsibility:

1. Patient autonomy — it is critically important to make sure that implementation of AI systems does not diminish the role of the patient in the decision-making process. The risks of excessive automatic trust of clinicians in AI advice and quality of informed consent should be taken into account.
2. Justice means preventing algorithmic discrimination and providing access to AI technologies to various groups. It is necessary to concentrate on data-centric risks and data representativeness.
3. Non-harming takes into account the possibility of delayed and systemic consequences associated with evolutionary changes and self-learning algorithms.
4. Transparency and explainability means ensuring interpretability of decisions and audit opportunities for both specialists and patients; overcoming the black box effect.
5. Mandatory ethical audit is analysis of compliance of artificial intelligence used with medical ethics, and regular revision of the risk matrix taking into account vulnerability of certain categories of patients and long-term consequences.

These provisions are shown in the matrix of foreseeable harm assessment and presented in details (see Table 2).

The use of the matrix allows to structure the process of identifying and minimizing predictable harm at all stages of AI-powered medical device life cycle, providing an integrated approach to risk management and compliance with regulatory and ethical requirements.

While developing medical AI systems, forming a culture of transparency is crucial for effective implementation of the predictable harm concept. This aspect includes open communication regarding technological limitations, active involvement of clinical specialists at all stages of product creation, and use of the safety through design principle involving integration of safety mechanisms directly into the system architecture. In contrast to existing regulatory approaches that focus primarily on technical characteristics and preliminary risk classification, the proposed concept assumes mandatory integration of ethical audits at each stage of the life cycle of a medical AI system. This requires multidisciplinary collaboration between developers, clinicians, ethicists, and patient community representatives to prevent algorithmic discrimination, preserve patient autonomy, and maintain equitable access to the benefits of technology. The matrix of predictable harm assessment, including ethical and clinical parameters, becomes not a simple documentation tool, but a platform for continuous dialogue between all participants in the process of introducing AI into clinical practice.

SUMMARY AND CONCLUSIONS

The conducted research allows us to formulate the following main conclusions:

1. The predictable harm concept is an effective tool for improving the safety of introducing artificial intelligence into medicine, which proactively identifies and minimizes risks.

Table 2. Matrix assessing predictable harm to medical AI systems

Stage	Key assessment issues	Tools	Responsible parties	Clinical aspects	Ethical aspects
Conceptualization and design	Compliance with the target application, coverage of clinical scenarios, technical feasibility	Ethical audit, analysis of clinical scenarios, review of the evidence base	Developers, clinical experts, ethics committees	Assessing clinical significance of the problem being solved, potential changes in clinical practice, and the risk/benefit ratio	Compliance with the values of the medical profession, ensuring patient autonomy, compliance with the principles of medical ethics
Development and training	Data representativeness, algorithm validity, resistance to extreme cases	Statistical analysis, algorithmic audit, simulation of boundary cases	Developers, data scientists, ML engineers	Coverage of diverse clinical situations, inclusion of rare cases, consideration of comorbid conditions	Prevention of discrimination and bias based on gender, age, ethnicity, socio-economic status
Validation and verification	Accuracy, specificity, sensitivity, robustness, productivity	Cross-validation, boundary case testing, external validation	Independent experts, clinicians, regulators	Assessing the impact on clinical decisions and treatment outcomes, comparison with the gold standard of diagnosis	Transparency and explainability of results, the possibility of challenging, protection from automated discrimination
Implementation and integration	Protocol compatibility, impact on clinical decisions, easy use	Simulation of working processes, testing in real conditions, audit of clinical pathways	Medical organizations, IT specialists, and clinicians	Assessing the impact on care provision process, decision-making time, integration into existing clinical protocols	Impact on doctor-patient relations, level of trust, preservation of clinical autonomy of the doctor
Post-marketing monitoring	Undesirable phenomena, production drift, unforeseen consequences	Real-world data analytics, incident reporting system, regular audit	Manufacturers, regulators, medical professionals, patient communities	Monitoring of deviations of clinical outcomes from expected ones, long-term impact on the quality of care, detection of rare complications	Considerating patients' experience, psychosocial consequences of AI use, assessment of the impact on medical care availability

2. The unique risks of using artificial intelligence require a differentiated management approach where integration of ethical aspects is mandatory.
3. According to the comparative analysis, the author's concept complements and expands existing regulatory approaches, providing a multidimensional, continuous and ethically supported harm assessment.
4. The prospects for further work are associated with universal methodologies and standardization of risk assessment practices for creation and application of artificial intelligence in healthcare.

Thus, it is essential to develop and implement the predictable harm concept in development and implementation of AI-powered medical devices as it can ensure an optimal balance between the innovative potential of these technologies and patient safety. A comparative analysis with existing regulatory approaches shows the advantages of the proposed concept in terms of multidimensional risk assessment and integration of ethical principles at all AI life cycle stages. The matrix of predictable harm, which includes parameters of clinical consequences and ethical assessment, allows us to proceed from formal risk management procedures to a systematic approach that takes into account both technological and humanitarian aspects of using artificial intelligence in healthcare. Promising trends of further research in this area include development of standardized risk assessment methodologies for various categories of AI systems, creation of validated ethical audit tools for medical AI solutions, and formation of unified regulatory requirements that synthesize technological standards with the principles of medical ethics and focus on long-term social consequences of introducing intelligent technologies in healthcare.

References

1. Grand View Research. Artificial Intelligence in Healthcare Market Size Report, 2021–2028. Available from URL: https://www.grandviewresearch.com/industry-analysis/artificial-intelligence-ai-healthcare-market (accessed: 01.05.2025).
2. Prikaz Ministerstva zdravookhraneniya Rossiyskoy Federatsii ot 7 iyulya 2020 g. № 686n «O vnesenii izmeneniy v prilozheniya № 1 i № 2 k prikazu Ministerstva zdravookhraneniya Rossiyskoy Federatsii ot 6 iyunya 2012 g. № 4n «Ob utverzhdenii nomenklaturnoy klassifikatsii meditsinskikh izdeliy» Ofitsial'nyy internet-portal zakona informatsii. Elektron. dan. Available from URL: https://www.pravo.gov.ru, svobodnyy. № publikatsii: 0001202008100015 (accessed: 10.05.2025). Russian.
3. Pis'mo Federal'noy sluzhby po nadzoru v sfere zdravookhraneniya ot 13 fevralya 2020 g. № 02I-297/20 «O programmnom obespechenii». [Tekst pis'ma opublikovan ne byl]. Russian.
4. «Software as a Medical Device»: Possible Framework for Risk Categorization and Corresponding Considerations. Available from URL: https://www.imdrf.org/sites/default/files/docs/imdrf/final/technical/imdrf-tech-140918-samd-framework-risk-categorization-141013.pdf (accessed: 10.05.2025).
5. Kodeks etiki v sfere II v meditsine i zdravookhranenii. Available from URL: https://ethics.a-ai.ru/ethics-of-medicine (accessed: 10.05.2025). Russian.
6. World Health Organization. Ethics and governance of artificial intelligence for health: WHO guidance. Available from URL: https://www.who.int/publications/i/item/9789240029200 (accessed: 01.05.2025).
7. European Parliament. Regulation (EU) 2017/745 of the European Parliament and of the Council of 5 April 2017 on medical devices, amending Directive 2001/83/EC, Regulation (EC) No 178/2002

and Regulation (EC) No 1223/2009 and repealing Council Directives 90/385/EEC and 93/42/EEC (Text with EEA relevance.). Available from URL: https://eur-lex.europa.eu/eli/reg/2017/745/oj/eng (accessed: 01.05.2025).
8. European Commission. Proposal for a Regulation of the European Parliament and of the Council laying down harmonised rules on Artificial Intelligence (Artificial Intelligence Act). Available from URL: https://digital-strategy.ec.europa.eu/en/library/proposal-regulation-laying-down-harmonised-rules-artificial-intelligence (accessed: 01.05.2025).
9. U. S. Food and Drug Administration. Artificial Intelligence and Machine Learning (AI/ML) Software as a Medical Device Action Plan. Available from URL: https://www.fda.gov/medical-devices/software-medical-device-samd/artificial-intelligence-and-machine-learning-software-medical-device (accessed: 01.05.2025).

Литература

1. Grand View Research. Artificial Intelligence in Healthcare Market Size Report, 2021–2028. Available from URL: https://www.grandviewresearch.com/industry-analysis/artificial-intelligence-ai-healthcare-market (accessed: 01.05.2025).
2. Приказ Министерства здравоохранения Российской Федерации от 7 июля 2020 г. № 686н «О внесении изменений в приложения № 1 и № 2 к приказу Министерства здравоохранения Российской Федерации от 6 июня 2012 г. № 4н «Об утверждении номенклатурной классификации медицинских изделий» Официальный интернет-портал правовой информации. Электрон. дан. Режим доступа: [Электронный ресурс] URL: https://www.pravo.gov.ru, свободный. № опубликования: 0001202008100015 (дата обращения: 10.05.2025).
3. Письмо Федеральной службы по надзору в сфере здравоохранения от 13 февраля 2020 г. № 02И-297/20 «О программном обеспечении». [Текст письма опубликован не был].
4. «Software as a Medical Device»: Possible Framework for Risk Categorization and Corresponding Considerations. Available from URL: https://www.imdrf.org/sites/default/files/docs/imdrf/final/technical/imdrf-tech-140918-samd-framework-risk-categorization-141013.pdf (accessed: 10.05.2025).
5. Кодекс этики в сфере ИИ в медицине и здравоохранении. Режим доступа: [Электронный ресурс] URL: https://ethics.a-ai.ru/ethics-of-medicine (accessed: 10.05.2025).
6. World Health Organization. Ethics and governance of artificial intelligence for health: WHO guidance. Available from URL: https://www.who.int/publications/i/item/9789240029200 (accessed: 01.05.2025).
7. European Parliament. Regulation (EU) 2017/745 of the European Parliament and of the Council of 5 April 2017 on medical devices, amending Directive 2001/83/EC, Regulation (EC) No 178/2002 and Regulation (EC) No 1223/2009 and repealing Council Directives 90/385/EEC and 93/42/EEC (Text with EEA relevance.). Available from URL: https://eur-lex.europa.eu/eli/reg/2017/745/oj/eng (accessed: 01.05.2025).
8. European Commission. Proposal for a Regulation of the European Parliament and of the Council laying down harmonised rules on Artificial Intelligence (Artificial Intelligence Act). Available from URL: https://digital-strategy.ec.europa.eu/en/library/proposal-regulation-laying-down-harmonised-rules-artificial-intelligence (accessed: 01.05.2025).
9. U. S. Food and Drug Administration. Artificial Intelligence and Machine Learning (AI/ML) Software as a Medical Device Action Plan. Available from URL: https://www.fda.gov/medical-devices/software-medical-device-samd/artificial-intelligence-and-machine-learning-software-medical-device (accessed: 01.05.2025).

CLASSIFICATION OF RISKS OF USING ARTIFICIAL INTELLIGENCE SYSTEMS IN THE FIELD OF MENTAL HEALTH

Semenova NV ✉, Martynyuk KL

Bekhterev National Medical Research Center of Psychiatry and Neurology, Saint Petersburg, Russia

The use of artificial intelligence systems (AIS) in healthcare is one of the promising solutions for improving access of citizens to modern medical technologies to prevent mental health disorders, stress-associated and psychosomatic diseases, to ensure early diagnosis and correction of mental disorders and their risk factors. However, an urgent problem is assessment of the risks of implementing AIS in the field of mental health and possibilities of managing such risks depending on various factors affecting clinical outcomes, as well as ethical aspects related to provision of this type of AIS-based medical care. The proposed risk classification system for the use of AIS in the field of mental health will expand the possibilities of introducing medical AIS into clinical practice while maintaining a high level of control over the risks of mental health disorders.

Keywords: mental health, psychological help, psychotherapy, artificial intelligence

Acknowledgements: the authors express their deep gratitude to Elena A. Volskaya, Candidate of Historical Sciences, Advisor to Director of N. A. Semashko National Research Institute of Public Health, Chairman of the Independent Interdisciplinary Committee for the Ethical Review of Clinical Trials, and Chairman of the Interuniversity Ethics Committee for supporting the idea of the article and advising on ethics issues.

Author contribution: Semenova NV — the idea of the article, problem statement, discussion of key substantive and ethical issues, planning and discussion of the article, editing the text of the article, its design; Martynyuk KL — participation in the discussion of the problem, study of world experience and literature on the topic, systematization and generalization of data, participation in discussion of the results, and writing the main text of the article.

Compliance with ethical standards: the ethics committee meeting was not held as the material for publication included theoretical provisions without involvement of patients; the article was prepared in compliance with ethical standards.

✉ **Correspondence should be addressed:** Natalia V. Semenova
Bekhtereva St., 3, St. Petersburg, 191019, Russia; org@bekhterev.ru

Received: 22.04.2025 **Accepted:** 20.05.2025 **Published online:** 08.06.2025

DOI: 10.24075/medet.2025.005

КЛАССИФИКАЦИЯ РИСКОВ ПРИМЕНЕНИЯ СИСТЕМ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА В СФЕРЕ ПСИХИЧЕСКОГО ЗДОРОВЬЯ

Н. В. Семенова ✉, К. Л. Мартынюк

Национальный медицинский исследовательский центр психиатрии и неврологии имени В. М. Бехтерева, Санкт-Петербург, Россия

Применение систем на основе искусственного интеллекта (СИИ) в здравоохранении является одним из перспективных решений для улучшения доступа граждан к современным медицинским технологиям профилактики нарушений психического здоровья, стресс-ассоциированных и психосоматических заболеваний, ранней диагностики и коррекции психических нарушений и факторов риска их развития. Однако актуальной проблемой является оценка рисков внедрения СИИ в сфере психического здоровья и возможности управления такими рисками в зависимости от различных факторов, влияющих на клинические исходы, а также этические аспекты, связанные с оказанием этого вида медицинской помощи с применением СИИ. Предложенная система классификации рисков применения СИИ в сфере психического здоровья позволит расширить возможности внедрения медицинских СИИ в клиническую практику при сохранении высокого уровня контроля рисков нарушения психического здоровья.

Ключевые слова: психическое здоровье, психологическая помощь, психотерапия, искусственный интеллект

Благодарности: авторы выражают глубокую признательность кандидату исторических наук, советнику директора ФГБНУ «Национальный НИИ общественного здоровья имени Н. А. Семашко», председателю Независимого междисциплинарного комитета по этической экспертизе клинических исследований, председателю Межвузовского комитета по этике Елене Алексеевне Вольской за поддержку идеи статьи и консультации по вопросам этики.

Вклад авторов: Н. В. Семенова — идея статьи, постановка проблемы, обсуждение ключевых содержательных и этических вопросов, планирование и обсуждение статьи, редактирование текста статьи, ее оформление; К. Л. Мартынюк — участие в обсуждении проблемы, изучение мирового опыта и литературы по теме, систематизация и обобщение данных, участие в обсуждении результатов, написание основного текста статьи.

Соблюдение этических стандартов: заседание этического комитета не проводилось, так как материалом для публикации послужили теоретические положения без участия пациентов; подготовка статьи проводилась с соблюдением этических норм.

✉ **Для корреспонденции:** Наталия Владимировна Семенова
ул. Бехтерева, д. 3, г. Санкт-Петербург, 191019, Россия; org@bekhterev.ru

Статья поступила: 22.04.2025 **Статья принята к печати:** 20.05.2025 **Опубликована онлайн:** 08.06.2025

DOI: 10.24075/medet.2025.005

High demand for medical, psychological and psychotherapeutic care among citizens, especially those living in metropolitan cities [1], which is accompanied by insufficient appealability of these types of medical aid in public health institutions due to persistent public stigma, contributes to stable inequality in access to advanced medical technologies for prevention of mental health disorders, stress-associated and psychosomatic diseases, and non-pharmacological intervention in mental disorders.

Late treatment leads to late care for a patient, usually with more pronounced disorders, increasing the requirements for

volume, complexity and cost of interventions with less potential for functioning restoration and quality of prognosis. It ultimately results in professional burnout of qualified personnel with increased disability and decreased productivity, deterioration of health and a general decrease in life quality of employees. Meanwhile, mental disorders (MD) have been the leading factor in the global burden of diseases for more than 20 years regarding the number of years of life with disability [2, 3] and maintaining a high burden on social funds.

The use of artificial intelligence systems (AIS) is one of the promising solutions for improving access of citizens to modern medical technologies for prevention of mental health disorders, stress-associated and psychosomatic diseases, early diagnosis and correction of mental disorders and risk factors for their development, as well as expanding the possibilities of psychotherapeutic interventions in case of mental disorders when medical, psychological and psychotherapeutic types of care are limitedly available due to the continuing shortage of appropriate specialized personnel. The first experience of its application is being actively discussed in scientific literature [4–12].

In recent years, researchers were optimistic about AI capabilities, as AI models are increasingly able to simulate a real interaction with humans. This can contribute not only to treatment delivery, but also to treatment adherence as compared to other forms of e-health. For example, conversational AI-based apps have proven their effectiveness in reducing symptoms of depression and anxiety, preventing stress, general distress, and negative affects, and improving well-being [4, 8, 13]. Meanwhile, an uncontrolled growth in the number of AI applications and difficulties in tracking the empirical results of using AIS are considered as significant negative factors [10–12].

The most discussed areas of AI application in the field of mental health include as follows:

- Supporting decisions taken during diagnosis of mental disorders and choosing the treatment strategy;
- The function of a “supervisor” and a “medical assistant” increasing commitment of people to preventive, diagnostic and therapeutic measures;
- Dynamic monitoring of the condition of patients suffering from mental disorders;
- Predicting the risk of mental disorder exacerbation;
- Non-drug control of symptoms of mental disorder using an “AI therapist”;
- Correction of emotional problems using AI models and methods of psychotherapy and psychocorrection;
- Prevention of emotional disorders, involving the use of AI, aimed at development of emotional intelligence and stress tolerance.

Taking into consideration high vulnerability of the citizens when they turn to specialists in mental health, limited practical use of technical devices in the clinical process by mental health professionals, and the emerging regulation of access of AI-based medical devices to circulation (according to Decree of the Government of the Russian Federation dated December 27, 2012 No. 1416 (as amended on November 24, 2020), and the fact that the AI-based MD belong to the 3rd (maximum) risk class) [14], it seems relevant to consider the risks of implementing AIS in the field of mental health and possibilities of managing such risks depending on various factors affecting clinical outcomes, as well as ethical aspects related to provision of this type of medical care using AIS.

The experience of medical use of AIS described in publications [2–18] helps us identify common sources of risks: those related to the technological features of development and operation of AI models and systems based on them; specific for use in the relevant clinical field; ethical and social aspects covering the basic principles of civil rights and freedom, fairness, confidentiality, security and transparency [15–18]. Meanwhile, the latter go through the entire MD life cycle: “The manufacturer must establish, document and maintain a continuous process of MD-associated identifying hazards, identifying and evaluating associated risks, managing these risks and monitoring such management throughout the MD life cycle (starting from design, including scientific research, and to decommissioning according to A.2.1 Scope of application) in accordance with the requirements of GOST ISO 14971” [2]. Manufacturers of medical AIS have even stricter obligations if they are guided by the Code of Ethics of Artificial Intelligence in National Healthcare [15]: “AIS developers must adhere to the ethical obligations and values followed by medical personnel in their actions towards patients in clinical practice, including the Code of Professional Ethics of a Doctor of the Russian Federation.” (Article 3 of the draft Code), which seems justified taking into account the specifics of AIS development and the inherent limitations of the ability to control individual risks during the operation of such systems in real clinical practice, especially when it is independently used by the patient.

Generalized sources of AI risks are given in PNS 840-2023 “Artificial Intelligence. An overview of ethical and social aspects” [17]. They include as follows:

- unauthorized means or methods of collecting, processing, or disclosing personal data;
- obtaining and using biased, inaccurate or unrepresentative data for AIS training;
- non-transparent machine learning (ML) decision-making or insufficient documentation, commonly referred to as lack of explainability;
- lack of tracking capability (iterative inaccuracy of AI models, working in an open contextual environment with unforeseen events and conditions);
- insufficient understanding of technology social impact after its introduction.

Specific sources of risks of using AIS in the field of mental health are as follows:

- Limited application of non-harm principle in case of incomplete fault tolerance, insufficient accuracy of operation or effectiveness of models for real clinical practice, or unpredictability of AIS when working under borderline conditions;
- Lack of transparency regarding the nature of the services provided by AIS and its representation as an assistant that uses therapeutic methods;
- Lack of control over the reviews and recommendations that users receive from offline AIS based on generative models;
- Limited understanding of clinical process components implemented by AIS, which form the basis for psychotherapeutic intervention effectiveness;
- AIS influence on autonomy and free will: formation of AIS-dependent response/behavior of patients due to over-accessibility, developing attachment (anthropomorphization) or excessive trust in AIS and loss of contact quality with a specialist, or development of anxiety, stress, and hypochondria due to AIS constant and frequent use;
- Increased resilience of the stigma associated with mental health disorders due to encouraging users to

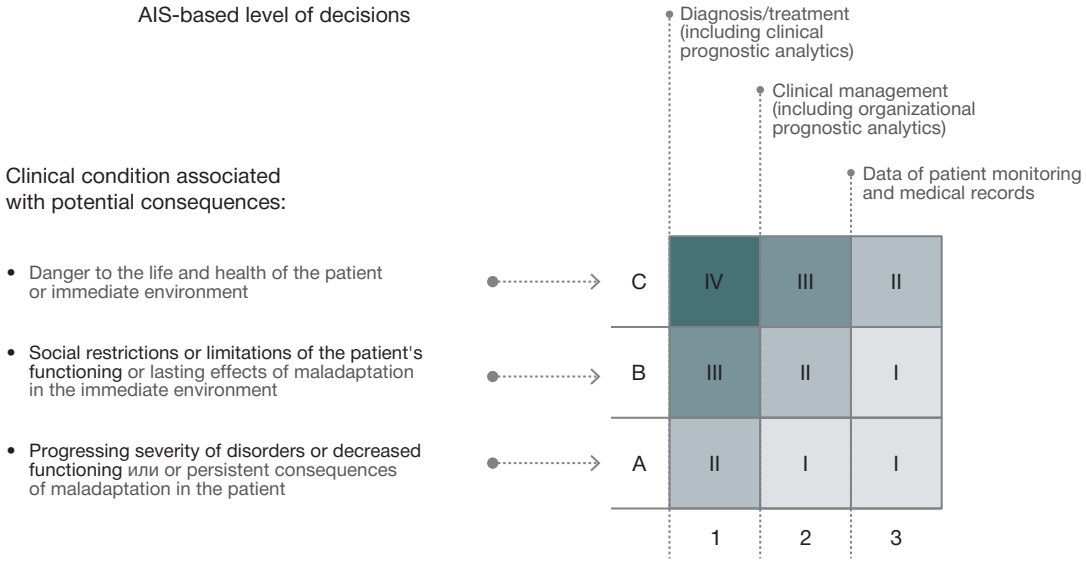

Fig. Classification of risks of using AIS in the field of mental health

use AIS or in the face of actual lack of alternatives, and expanding the possibilities of risky self-treatment when the system response is interpreted in an incorrect way.

Thus, to model the risks of using AIS specific to the field of mental health, it is necessary to determine the semantic space in the following areas: the goals of using AIS, continuum of relevant clinical conditions and outcomes, safety of software packages, and ethical certainty.

Currently, when technologies in the field of mental health are being developed, AIS are applicable for the following purposes (based on GOST R 59525-2021 Health informatics. Intelligent methods of medical data processing. Main provisions" [19–20]):

– to diagnose, prevent, observe, treat or relieve disorders;
– to support the vital activity (if the patient's autonomy is included into the term as well);
– functioning of medical decision support systems;
– predicting the appearance and/or development of diseases based on genetic data.

Meanwhile, the last point in the field of mental health can also be interpreted based on bio-psycho-social factors. Thus, the main patterns of reaction/behavior, as a rule, are unknowingly borrowed by the child from his family and form a characteristic pattern, which further becomes a part of the child's personality, and can no less affect resistance to stress and likelihood of developing stress-related diseases.

Consequently, the use of AIS is justified in accordance with the principles of evidence-based medicine. It is limited to conditions ranging from subclinical disorders recognized by diagnostic methods validated for the Russian population to threats to life and health of the patient or people around the patient (the latter is relevant for a number of chronic and long-term mental disorders with severe persistent or frequently aggravated painful manifestations).

Taking into account the principle of classification of the software (including AIS) used while providing medical care, in terms of safety, a class is assigned according to the risk of harm to the patient, user or other persons, based on a dangerous situation to which the program system (PS) can contribute in the worst-case scenario [2]:

– Class A: PS can contribute to a dangerous situation that does not lead to an unacceptable risk;
– Class B: PS can contribute to a dangerous situation that leads to an unacceptable risk, and the resulting possible harm is not a serious injury;
– Class C: PS can contribute to a dangerous situation that leads to an unacceptable risk, and as a result, possible harm can include death or serious injury.

It should be clarified that in case of mental health disorders, the concept of an unacceptable risk that does not result in a serious injury or death is most consistent with consequences in the form of social restrictions or limited functioning of patients, as well as the persistent consequences of maladaptation associated with close interaction with such a patient in the immediate environment (cohabiting relatives, especially minors in the process of personality formation, and persons who have been informal caregivers for a long time). The boundary condition for gradation within the acceptable risk range should include progression of the severity of disorders or stable consequences of maladaptation in the patient, since management of such risks is fully available within the framework of a high-quality clinical process (adherence to treatment, therapeutic alliance, compliance with clinical recommendations, scientifically based innovative methods of intervention).

To expand the possibilities of AI implementation in supporting and auxiliary processes, it is advisable to take into account the importance of information for making medical decisions, which is processed by AIS, and to varying degrees may affect the level of clinical risks [18] in the range of:

– data (including AIS interpreted data) for diagnosis or treatment (including clinical predictive analytics);
– data for clinical management (including organizational predictive analytics);
– data of patient monitoring and medical records (including those entered into AIS by the patient or caregiver).

It is acceptable to reduce the risk for each step of the above gradation of information significance, as this corresponds to a reduction in the impact of data on clinical decisions, due to the possibility of maximum control of individual risks within the framework of a qualitative clinical process.

Thus, it is possible to propose the following gradation of the risk of using AIS in the field of mental health n the form of a 2-dimensional matrix (Fig.):

A) regarding safety (clinical condition associated with potential consequences):

– Risk category IV — danger to the life and health of the patient or immediate environment;
– Risk category III — social restrictions or limitations of the patient's functioning or lasting effects of maladaptation in the immediate environment;
– Risk category II is the progressing severity of disorders or persistent consequences of maladaptation in the patient;

B) regarding the process significance (the level of decisions based on the AIS provided information):
– Risk category N — diagnosis or treatment (including clinical predictive analytics);
– Risk category N-1 — clinical management (including organizational predictive analytics);
– Risk category N-2 ≥ 1 — data of patient monitoring and medical records.

The proposed risk classification system for using AIS in the field of mental health will expand the possibilities of introducing medical AIS into clinical practice while maintaining a high level of control over both individual and public (micro- and macrosocial) risks of mental health disorders and development of stress-associated diseases, as well as increasing availability of qualified care and, accordingly, earlier interventions for mental disorders with a high potential for restoring functioning and maintaining the quality of life of patients. At the same time, it provides the possibility of high-quality ethical risk management for the specialized use of AIS throughout the entire life cycle (from design, including scientific research, post-registration monitoring of quality/effectiveness and recalibration of AI models, to decommissioning) of an AI-based medical device at all levels of regulatory and industry requirements.

References

1. Hołownia-Voloskova M, Tarbastaev A, Golicki D. Population norms of health-related quality of life in Moscow, Russia: the EQ-5D-5L-based survey. Quality of Life Research. 2021; 30(3): 831–840.
2. Global Burden of Disease Study 2019 (GBD 2019) Available from URL: https://ghdx.healthdata.org/gbd-2019 (accessed: 11.03.2025).
3. Vos T, Lim SS, Abbafati C, Abbas KM, Abbasi M, Abbasifard M, et al. Global Burden of 369 Diseases and Injuries in 204 Countries and Territories, 1990–2019: A Systematic Analysis for the Global Burden of Disease Study 2019. The Lancet. 2020; 396.
4. Espejo G, Reiner W, Wenzinger M. Exploring the Role of Artificial Intelligence in Mental Healthcare: Progress, Pitfalls, and Promises. Cureus. 2023; 15(9): e44748.
5. Groot J, MacLellan A, Butler M, Todor E, Zulfiqar M, Thackrah T, Clarke C, Brosnan M, Ainsworth B. The Effectiveness of Fully Automated Digital Interventions in Promoting Mental Well-Being in the General Population: Systematic Review and Meta-Analysis. JMIR Ment Health. 2023; 10: e44658.
6. Grote T, Berens P. A paradigm shift? — On the ethics of medical large language models. Bioethics. 2024; 38: 383–390.
7. Jain G, Pareek S, Carlbring P. Revealing the source: How awareness alters perceptions of AI and human-generated mental health responses. Internet Interv. 2024; 36: 100745.
8. Lee EE, Torous J, De Choudhury M, Depp CA, Graham SA, Kim HC, Paulus MP, Krystal JH, Jeste DV. Artificial Intelligence for Mental Health Care: Clinical Applications, Barriers, Facilitators, and Artificial Wisdom. Biol Psychiatry Cogn Neurosci Neuroimaging. 2021; 6(9).
9. Rahsepar Meadi M, Bernstein JS, Batelaan N, van Balkom AJL, Metselaar S. Does a lack of emotions make chatbots unfit to be psychotherapists? Bioethics. 2024; 1–8.
10. Rogan J, Bucci S, Firth J. Health Care Professionals' Views on the Use of Passive Sensing, AI, and Machine Learning in Mental Health Care: Systematic Review With Meta-Synthesis. JMIR Ment Health. 2024; 11: e49577.
11. Tornero-Costa R, Martinez-Millana A, Azzopardi-Muscat N, Lazeri L, Traver V, Novillo-Ortiz D. Methodological and Quality Flaws in the Use of Artificial Intelligence in Mental Health Research: Systematic Review. JMIR Ment Health. 2023; 10: e42045.
12. Vandersluis R, Savulescu J. The selective deployment of AI in healthcare. Bioethics. 2024; 38: 391–400.
13. He Y, Yang L, Qian C, Li T, Su Z, Zhang Q, Hou X. Conversational Agent Interventions for Mental Health Problems: Systematic Review and Meta-analysis of Randomized Controlled Trials. J Med Internet Res. 2023; 2: e43862.
14. Postanovlenie Pravitel'stva RF ot 27.12.2012 № 1416 (s izmenenijami na 24.11.2020) «Ob utverzhdenii Pravil gosudarstvennoj registracii medicinskih izdelij». Available from URL: http://government.ru/docs/all/85645/ (accessed: 12.03.2025). Russian.
15. AI'jans v sfere iskusstvennogo intellekta. Kodeks jetiki v sfere iskusstvennogo intellekta. Available from URL: https://ethics.a-ai.ru/ (accessed: 11.03.2025). Russian.
16. Nacional'nyj centr razvitija iskusstvennogo intellekta pri Pravitel'stve Rossijskoj Federacii. Informacionno-analiticheskaja spravka «Zarubezhnaja praktika ustanovlenija trebovanij k bezopasnosti II» (2023). Available from URL: https://ai.gov.ru/ (accessed: 10.03.2025). Russian.
17. PNST 840-2023. Iskusstvennyj intellekt. Obzor jeticheskih i obshhestvennyh aspektov. Available from URL: https://nd.gostinfo.ru/document/7475760.aspx (accessed: 11.03.2025). Russian.
18. International Medical Device Regulators Forum. Software as a Medical Device: Framework for Risk Categorization and Corresponding Considerations. Available from URL: https://www.imdrf.org/sites/default/files/docs/imdrf/final/technical/imdrf-tech-140918-samd-framework-risk-categorization-141013.pdf (accessed: 10.03.2025).
19. GOST ISO 14971-2021. Izdelija medicinskie. Primenenie menedzhmenta riska k medicinskim izdelijam. Available from URL: https://nd.gostinfo.ru/document/5307378.aspx (accessed: 11.03.2025). Russian.
20. GOST R 59525-2021. Informatizacija zdorov'ja. Intellektual'nye metody obrabotki medicinskih dannyh. Osnovnye polozhenija. Available from URL: https://nd.gostinfo.ru/document/6856334.aspx (accessed: 11.03.2025). Russian.

Литература

1. Hołownia-Voloskova M, Tarbastaev A, Golicki D. Population norms of health-related quality of life in Moscow, Russia: the EQ-5D-5L-based survey. Quality of Life Research. 2021; 30 (3): 831–840.
2. Global Burden of Disease Study 2019 (GBD 2019) Available from URL: https://ghdx.healthdata.org/gbd-2019 (accessed: 11.03.2025).
3. Vos T, Lim SS, Abbafati C, Abbas KM, Abbasi M, Abbasifard M, et al. Global Burden of 369 Diseases and Injuries in 204 Countries and Territories, 1990–2019: A Systematic Analysis for the Global Burden of Disease Study 2019. The Lancet. 2020; 396.
4. Espejo G, Reiner W, Wenzinger M. Exploring the Role of Artificial Intelligence in Mental Healthcare: Progress, Pitfalls, and Promises. Cureus. 2023; 15(9): e44748.
5. Groot J, MacLellan A, Butler M, Todor E, Zulfiqar M, Thackrah T, Clarke C, Brosnan M, Ainsworth B. The Effectiveness of Fully Automated Digital Interventions in Promoting Mental Well-Being

in the General Population: Systematic Review and Meta-Analysis. JMIR Ment Health. 2023; 10: e44658.
6. Grote T, Berens P. A paradigm shift? — On the ethics of medical large language models. Bioethics. 2024; 38: 383–390.
7. Jain G, Pareek S, Carlbring P. Revealing the source: How awareness alters perceptions of AI and human-generated mental health responses. Internet Interv. 2024; 36:100745.
8. Lee EE, Torous J, De Choudhury M, Depp CA, Graham SA, Kim HC, Paulus MP, Krystal JH, Jeste DV. Artificial Intelligence for Mental Health Care: Clinical Applications, Barriers, Facilitators, and Artificial Wisdom. Biol Psychiatry Cogn Neurosci Neuroimaging. 2021; 6(9).
9. Rahsepar Meadi M, Bernstein JS, Batelaan N, van Balkom AJL, Metselaar S. Does a lack of emotions make chatbots unfit to be psychotherapists? Bioethics. 2024; 1–8.
10. Rogan J, Bucci S, Firth J. Health Care Professionals' Views on the Use of Passive Sensing, AI, and Machine Learning in Mental Health Care: Systematic Review With Meta-Synthesis. JMIR Ment Health. 2024; 11: e49577.
11. Tornero-Costa R, Martinez-Millana A, Azzopardi-Muscat N, Lazeri L, Traver V, Novillo-Ortiz D. Methodological and Quality Flaws in the Use of Artificial Intelligence in Mental Health Research: Systematic Review. JMIR Ment Health. 2023; 10: e42045.
12. Vandersluis R, Savulescu J. The selective deployment of AI in healthcare. Bioethics. 2024; 38: 391–400.
13. He Y, Yang L, Qian C, Li T, Su Z, Zhang Q, Hou X. Conversational Agent Interventions for Mental Health Problems: Systematic Review and Meta-analysis of Randomized Controlled Trials. J Med Internet Res. 2023; 25: e43862.
14. Постановление Правительства РФ от 27.12.2012 № 1416 (с изменениями на 24.11.2020) «Об утверждении Правил государственной регистрации медицинских изделий». Режим доступа: [Электронный ресурс] URL: http://government.ru/docs/all/85645/ (дата обращения: 11.03.2025).
15. Альянс в сфере искусственного интеллекта. Кодекс этики в сфере искусственного интеллекта. Режим доступа: [Электронный ресурс] URL: https://ethics.a-ai.ru/ (дата обращения: 11.03.2025).
16. Национальный центр развития искусственного интеллекта при Правительстве Российской Федерации. Информационно-аналитическая справка «Зарубежная практика установления требований к безопасности ИИ» (2023). Режим доступа: [Электронный ресурс] URL: https://ai.gov.ru/ (дата обращения: 11.03.2025).
17. ПНСТ 840-2023. Искусственный интеллект. Обзор этических и общественных аспектов. Режим доступа: [Электронный ресурс] URL: https://nd.gostinfo.ru/document/7475760.aspx (дата обращения: 11.03.2025).
18. International Medical Device Regulators Forum. Software as a Medical Device: Framework for Risk Categorization and Corresponding Considerations. Режим доступа: [Электронный ресурс] URL: https://www.imdrf.org/sites/default/files/docs/imdrf/final/technical/imdrf-tech-140918-samd-framework-risk-categorization-141013.pdf (дата обращения: 11.03.2025).
19. ГОСТ ISO 14971-2021. Изделия медицинские. Применение менеджмента риска к медицинским изделиям. Режим доступа: [Электронный ресурс] URL: https://nd.gostinfo.ru/document/5307378.aspx (дата обращения: 11.03.2025).
20. ГОСТ Р 59525-2021. Информатизация здоровья. Интеллектуальные методы обработки медицинских данных. Основные положения. Режим доступа: [Электронный ресурс] URL: https://nd.gostinfo.ru/document/6856334.aspx (дата обращения: 11.03.2025).

NEUROETHICAL REGULATION OF PEDIATRIC DEEP BRAIN STIMULATION (DBS) IN THE RUSSIAN FEDERATION: RISKS OF UNJUSTIFIED USE

Klyuyeva PA ✉

Yaroslavl State Medical University, Yaroslavl, Russia

The study covers the problem of insufficient regulation of pediatric deep brain stimulation (DBS) in Russia, which causes neuroethical dilemmas and risk of unjustified expansion of indications. The study is relevant because of growing use of DBS in children with severe neurological disorders and lack of adapted standards. The aim of the study is to identify gaps in regulatory system and develop recommendations for ethical and clinical regulation of DBS in children. The current clinical recommendations of the Ministry of Health of the Russian Federation, international consensuses and protocols of leading Russian centers were used as a basis for the research. The main methods included analysis of regulatory documents, comparative and critical analysis of existing standards and ethical approaches. The results show that Russian practice lacks age-specific standards and assessment algorithms despite the regulations. The risks of expanded indications and pressure on patients and their families, which may lead to unregulated experimentation, have been identified. As a conclusion, the need to develop specialized recommendations and strengthen ethical standards to ensure safety and effectiveness of DBS use in children is proposed.

Keywords: neuroethics, DBS, dystonia, cerebral palsy, Tourette's syndrome, pediatrics, neurosurgery

✉ **Correspondence should be addressed:** Polina A. Klyuyeva
44 Karl Marx Street, Gavrilov-Yam, Yaroslavl region, Russia; klyuyeva_polina@mail.ru

Received: 06.06.2025 **Accepted:** 20.06.2025 **Published online:** 30.06.2025

DOI: 10.24075/medet.2025.010

НЕЙРОЭТИЧЕСКОЕ РЕГУЛИРОВАНИЕ ДЕТСКОЙ DBS В РФ: РИСКИ НЕОБОСНОВАННОГО ПРИМЕНЕНИЯ

П. А. Клюева ✉

Ярославский государственный медицинский университет, Ярославль, Россия

Проблема исследования заключается в недостаточной регуляции применения глубокой стимуляции мозга у детей в России, что вызывает нейроэтические дилеммы и риск необоснованного расширения показаний. Актуальность обусловлена ростом использования DBS у педиатрической группы с тяжелыми неврологическими расстройствами и отсутствием адаптированных стандартов. Цель исследования — выявить пробелы в нормативной базе и разработать рекомендации по этическому и клиническому регулированию применения DBS у детей. В качестве материала использованы действующие клинические рекомендации Минздрава России, международные консенсусы и протоколы ведущих российских центров. Основные методы включали анализ нормативных документов, сравнительный и критический анализ существующих стандартов и этических подходов. Результаты показали, что российская практика страдает отсутствием возрастных стандартов и алгоритмов оценки, несмотря на наличие регламентов. Обнаружены риски расширения показаний и давления на пациентов и их семьи, что может привести к нерегулируемому экспериментированию. В качестве вывода предложена необходимость разработки специализированных рекомендаций и усиления этических стандартов для обеспечения безопасности и эффективности применения DBS у детей.

Ключевые слова: нейроэтика, DBS, дистония, детский церебральный паралич, синдром Туретта, педиатрия, нейрохирургия

✉ **Для корреспонденции:** Полина Анатольевна Клюева
ул. Карла Маркса, д. 44, г. Гаврилов-Ям, Ярославская обл., Россия; klyuyeva_polina@mail.ru

Статья поступила: 06.06.2025 **Статья принята к печати:** 20.06.2025 **Опубликована онлайн:** 30.06.2025

DOI: 10.24075/medet.2025.010

INTRODUCTION

Deep brain stimulation (DBS), which has demonstrated its effectiveness in severe neurological disorders in adults (Parkinson's disease, essential tremor), is increasingly being introduced into pediatric practice. When used in children with drug-resistant forms of dystonia (DYT1, PKAN), epilepsy (Lennox-Gastaut syndrome), Tourette's syndrome, and cerebral palsy, a significant improvement in motor functions and a decreased frequency of seizures was found [1, 2]. However, pediatric treatment has certain specifics due to continued development of the brain, impossibility of obtaining a full informed consent, and high vulnerability, giving rise to a complex of serious neuroethical dilemmas that go beyond medical risks.

The purpose of this study is to conduct a critical neuroethical analysis of the existing regulatory system of using DBS in children in the Russian Federation (based on current regulatory documents and clinical practices). It is done to identify potential risks of unjustified expansion of indications, pressure on patients and their families, and use of off-label methods that can turn therapeutic intervention into unregulated experimentation.

MATERIALS AND METHODS

While preparing for this study, the following categories of documents were analyzed.

1. Current clinical recommendations of the Ministry of Health of the Russian Federation (2023–2025) on nosologies involving pediatric use of DBS (dystonia G24, epilepsy G40, cerebral palsy G80, Tourette syndrome F95.2) [3, 4].
2. International Consensus on Neurostimulation in Children (CAPSIT-PD): CAPSIT-PD (Core Assessment Program for Surgical Interventional Therapies in Parkinson's Disease):

– Designed to standardize assessment of patients with Parkinson's disease before neurosurgical interventions, including DBS. However, the protocol is aimed at adults only and not adapted for children. It includes neuropsychological testing, assessment of motor functions (UPDRS) and quality of life, but its use in children is limited due to differences in pathogenesis and age-related cognitive features [5, 6].
– Critical limitation: a study performed in 2015 showed that only 40% of adult patients could absolutely tolerate preoperative CAPSIT-PD testing due to fatigue and complexity of tasks. In children, these risks are increased multiple times [6, 7].

Protocols of ethical committees of the leading neurosurgical centers of the Russian Federation:

1) Bekhterev Psychoneurological Research Institute (Saint Petersburg)
– Two-level informed consent: Signature of parents and written consent of a child who is 14 years or older. Mandatory inclusion of paragraphs about the risks of cognitive impairment and irreversibility of stimulation effects.
– Peer review: for children with psychiatric comorbidities (for example, ASD), the decision is made by a council consisting of a neurosurgeon, a pediatric psychiatrist and a bioethicist.
2) Burdenko National Medical Research Center for Neurosurgery (Moscow) — Off-label use restriction: prohibition of DBS for unapproved indications (for example, autism without autoaggression) without approval of the central ethics committee.

Analysis of the existing regulatory system, including current clinical guidelines of the Ministry of Health of the Russian Federation, international consensus (CAPSIT-PD) and protocols of the ethics committees from leading centers revealed that there are no sufficient validated and directly adapted data for pediatric practice of deep brain stimulation despite availability of regulatory documents applicable to DBS in general. The paradox is that formal availability of structured approaches (including requirements for informed consent and ethical assessment) presents a contrast to the acute shortage of specific, age-sensitive algorithms for patient selection, preoperative assessment and prediction of outcomes in children.

STUDY RESULTS

Analysis of the regulatory documents of the Ministry of Health of the Russian Federation (2023–2025), international consensuses (CAPSIT-PD) and protocols of the ethical committees of leading neurosurgical centers in Russia have shown that the existing system for regulating the use of deep brain stimulation (DBS) in children has significant gaps. Despite the formal existence of regulatory acts and requirements for informed consent, there are no adequately adapted clinical guidelines for pediatrics and standardized algorithms for patient selection, preoperative assessment and prediction of outcomes.

The international CAPSIT-PD protocol, which is aimed at adult patients with Parkinson's disease, does not take into account age-specific characteristics and cognitive differences in children. This limits its applicability in pediatric practice. Additionally, ethical difficulties related to obtaining informed consent and need for a comprehensive assessment of patients with psychiatric comorbidities have been identified.

Protocols of the leading Russian centers provide for collegial decision-making and restriction of off-label use of DBS. However, practice shows the risk of unjustified expansion of indications and pressure on patients and families. As a result, there is a lack of clear regulatory and ethical guidelines, which can lead to unregulated experiments with treatment of children suffering from severe neurological disorders.

Thus, the results emphasize the need to develop specialized, age-appropriate clinical guidelines and ethical standards to use DBS in pediatric patients in Russia, taking into account neuroethical aspects and protection of the rights of minors.

DISCUSSION OF THE RESULTS

A study of pediatric DBS regulation in the Russian Federation showed as follows:

1. The need for pediatric adaptation: current clinical guidelines contain general provisions on DBS, but detailed selection and evaluation algorithms that take into account the child's development have not been fully developed yet.
2. CAPSIT-PD limitations: use of this protocol in pediatrics is hampered because its focuses on adults with Parkinson's disease and the tests are too complex for children [6, 7].
3. Ethical challenges: existing approaches to informed consent (including consent of adolescents) require further development of methods that provide a deep understanding of long-term aspects of child and family treatment.

These observations are consistent with international experience where pediatric DBS standards are required and ethical approaches have to be clarified [8, 9].

We also identified some factors influencing the situation:

Relatively recent introduction of DBS in pediatrics, objective difficulty of creating universal standards for the developing brain, and need in additional resources.

CONCLUSIONS

The study revealed a gap between a formal regulation and a real shortage of adequate pediatric instruments and standards for DBS in the Russian Federation. Elimination of this gap through development of specialized recommendations, adapted assessment protocols, and strengthening the neuroethical component is a top priority that ensures safety, effectiveness, and ethics of using this method in children.

References

1. A systematic review of deep brain stimulation for the treatment of drug-resistant epilepsy in childhood. PubMed. Available from URL: https://pubmed.ncbi.nlm.nih.gov/30544364/ (accessed: 25.04.2025). Russian.
2. The use of DBS in the treatment of childhood cerebral palsy: a systematic review. PubMed. Available from URL: https://pubmed.ncbi.nlm.nih.gov/40407856/ (accessed: 25.04.2025). Russian.
3. Clinical recommendations: epilepsy and epileptic status in adults and children. Section 3.2. Available from URL: https://vk.com/doc453806649_687832800?hash=j0L4muvXTju2Ji0PcunzFJU6S2mSUY3dSLNp9zfEO60&dl=uArMIeSHlGo76IFzqOM6sP4ueK

pCZHTlGptFrZZzyTD&from_module=vkmsg_desktop (accessed: 25.04.2025). Russian.
4. Dystonia. In the section: Clinical and rheumatological aspects. Section 3.2.1 Available from URL: https://vk.com/doc453806649_687832864?hash=uBufhc7jHW0fMO9a88E8iaryB1qUkFwvxMlL7qQmq8z&dl=zMBQLfDrwOSpuliaCjTgu0YLcuvAEH2iVo7rSmZDjd4&from_module=vkmsg_desktop (accessed: 25.04.2025). Russian.
5. Core assessment program for surgical interventional therapies in Parkinson's disease (CAPSIT-PD). PubMed. Available from URL: https://pubmed.ncbi.nlm.nih.gov/10435493/ (accessed: 25.04.2025). Russian.
6. The Core Assessment Program for Surgical Interventional Therapies in Parkinson's Disease (CAPSIT-PD): Tolerability of Preoperative Neuropsychological Testing for Deep Brain Stimulation in Parkinson's Disease. PubMed. Available from URL: https://pubmed.ncbi.nlm.nih.gov/30363547/ (accessed: 25.04.2025). Russian.
7. Deep brain stimulation in children: experience and technical pearls. PubMed. Available from URL: https://pubmed.ncbi.nlm.nih.gov/22132914/ (accessed: 25.04.2025). Russian.
8. Deep Brain Stimulation for Pediatric Dystonia: A review of the literature and suggested programming algorithm. PMC. Available from URL: https://pmc.ncbi.nlm.nih.gov/articles/PMC9912476/ (accessed: 25.04.2025). Russian.
9. Deep Brain Stimulation for Pediatric Dystonia. ScienceDirect. Available from URL: https://www.sciencedirect.com/science/article/abs/pii/S1071909121000243/ (accessed: 25.04.2025). Russian.

Литература

1. Систематический обзор глубокой стимуляции мозга для лечения лекарственно-устойчивой эпилепсии у детей. PubMed. Режим доступа: [Электронный ресурс] URL: https://pubmed.ncbi.nlm.nih.gov/30544364/ (дата обращения: 25.04.2025).
2. Применение DBS в лечении детского церебрального паралича: систематический обзор. PubMed. Режим доступа: [Электронный ресурс] URL: https://pubmed.ncbi.nlm.nih.gov/40407856/ (дата обращения: 25.04.2025).
3. Клинические рекомендации: эпилепсия и эпилептический статус у взрослых и детей. Раздел 3.2. Режим доступа: [Электронный ресурс] URL: https://vk.com/doc453806649_687832800?hash=j0L4muvXTju2Ji0PcunzFJU6S2mSUY3dSLNp9zfEO60&dl=uArMleSHlGo76lFzqOM6sP4ueKpCZHTlGptFrZZzyTD&from_module=vkmsg_desktop (дата обращения: 25.04.2025).
4. Дистония. В разделе: Клинико-ревматологические аспекты. Раздел 3.2.1. Режим доступа: [Электронный ресурс] URL: https://vk.com/doc453806649_687832864?hash=uBufhc7jHW0fMO9a88E8iaryB1qUkFwvxMlL7qQmq8z&dl=zMBQLfDrwOSpuliaCjTgu0YLcuvAEH2iVo7rSmZDjd4&from_module=vkmsg_desktop (дата обращения: 25.04.2025).
5. Основная программа оценки хирургических методов лечения болезни Паркинсона (CAPSIT-PD). PubMed. Режим доступа: [Электронный ресурс] URL: https://pubmed.ncbi.nlm.nih.gov/10435493/ (дата обращения: 25.04.2025).
6. Основная программа оценки хирургических методов лечения болезни Паркинсона (CAPSIT-PD): переносимость предоперационного нейропсихологического тестирования для глубокой стимуляции мозга при болезни Паркинсона. PubMed. Режим доступа: [Электронный ресурс] URL: https://pubmed.ncbi.nlm.nih.gov/30363547/ (дата обращения: 25.04.2025).
7. Глубокая стимуляция мозга у детей: опыт и технические тонкости. PubMed. Режим доступа: [Электронный ресурс] URL: https://pubmed.ncbi.nlm.nih.gov/22132914/ (дата обращения: 25.04.2025).
8. Глубокая стимуляция мозга при детской дистонии: обзор литературы и предлагаемый алгоритм программирования. PMC. Режим доступа: [Электронный ресурс] URL: https://pmc.ncbi.nlm.nih.gov/articles/PMC9912476/ (дата обращения: 25.04.2025).
9. Глубокая стимуляция мозга при детской дистонии. ScienceDirect. Режим доступа: [Электронный ресурс] URL: https://www.sciencedirect.com/science/article/abs/pii/S1071909121000243/ (дата обращения: 25.04.2025).

LARGE LANGUAGE MODELS IN MEDICINE: CURRENT ETHICAL CHALLENGES

Kostrov SA ✉, Potapov MP

Yaroslavl State Medical University, Yaroslavl, Russia

The article analyzes the latest ethical challenges associated with introduction of large language models (LLMs) in medicine and healthcare. Various LLM architectures, stages of their training (pretraining, pretuning, reinforcement learning from human feedback) and criteria for quality of training data are reviewed. The emphasis is on a range of ethical issues such as copyright for AI-generated content; systematic bias in algorithms and risk of generating false information; a need to ensure transparency and explainability of AI (XAI); issues of confidentiality and protection of personal medical data, including difficulties with anonymization and obtaining informed consent. Aspects of legal responsibility for using LLMs in clinical practice are also analyzed and technological solutions (federated learning, homomorphic encryption) to minimize risks are discussed. The need for an integrated approach combining technological improvement, development of ethical standards, adaptation of legislation and critical supervision of the medical community is emphasized to ensure safe and effective integration of LLMs into clinical practice.

Keywords: artificial intelligence in medicine, large language models, generative text authorship, explainable AI, federated learning AI, bias, cybersecurity

Author contribution: Potapov MP — research planning, analysis, editing; Kostrov SA — collection, analysis, interpretation of data, preparation of a draft manuscript.

✉ Correspondence should be addressed: Sergey A. Kostrov
Revolutsionnaya str., 5., Yaroslavl, 150000, Russia; kosea@ysmu.ru

Received: 06.05.2025 Accepted: 20.05.2025 Published online: 29.06.2025

DOI: 10.24075/medet.2025.008

БОЛЬШИЕ ЯЗЫКОВЫЕ МОДЕЛИ В МЕДИЦИНЕ: АКТУАЛЬНЫЕ ЭТИЧЕСКИЕ ВЫЗОВЫ

С. А. Костров ✉, М. П. Потапов

Ярославский государственный медицинский университет, Ярославль, Россия

Статья посвящена анализу актуальных этических вызовов, связанных с внедрением больших языковых моделей (LLM) в сферу медицины и здравоохранения. Рассматриваются различные архитектуры LLM, этапы их обучения (предобучение, донастройка, обучение с подкреплением на основе обратной связи от человека) и критерии качества обучающих данных. Основное внимание уделяется комплексу этических проблем: вопросам авторского права на контент, сгенерированный искусственным интеллектом (ИИ); систематической предвзятости алгоритмов и риску генерации недостоверной информации; необходимости обеспечения прозрачности и объяснимости ИИ (XAI); проблемам конфиденциальности и защиты персональных медицинских данных, включая сложности анонимизации и получения информированного согласия. Также анализируются аспекты юридической ответственности за применение LLM в клинической практике и обсуждаются технологические решения (федеративное обучение, гомоморфное шифрование) для минимизации рисков. Подчеркивается необходимость комплексного подхода, сочетающего технологическое совершенствование, разработку этических стандартов, адаптацию законодательства и критический надзор медицинского сообщества для безопасной и эффективной интеграции LLM в клиническую практику.

Ключевые слова: искусственный интеллект в медицине, большие языковые модели, авторское право генеративного текста, объяснимый ИИ, федеративное обучение ИИ, предвзятость, кибербезопасность

Вклад авторов: М. П. Потапов — планирование исследования, анализ, редактирование; С. А. Костров — сбор, анализ, интерпретация данных, подготовка черновика рукописи.

✉ Для корреспонденции: Сергей Александрович Костров
ул. Революцилнная, д. 5, г. Ярославль, 150000, Россия; kosea@ysmu.ru

Статья поступила: 06.05.2025 Статья принята к печати: 20.05.2025 Опубликована онлайн: 29.06.2025

DOI: 10.24075/medet.2025.008

Over the past five years, artificial intelligence (AI) has become one of the fundamental technologies launching transformation of the basic paradigms of medicine and healthcare system [1, 2]. Recognizing potentially inflated expectations associated with this technology, it is necessary to clarify the terminology used below. In scientific and professional discourse, it is natural to distinguish between two main concepts of AI. The first one is artificial general intelligence (AGI), also known as strong artificial intelligence (AI), a hypothetical form of AI that can learn universally and solve problems like a human, which is only theoretical and has not been implemented in practice yet; the second one is artificial narrow intelligence (ANI), also referred to as weak AI, an existing software system that helps a person solve specific, clearly limited tasks, such as diagnosing diseases using medical images or automatization of routine operational processes.

The general term AI denotes ANI, which is used in practical medicine today.

Two major classes of weak AI are distinguished: descriptive and generative AI. Descriptive systems analyze and interpret data (including numerical, textual, graphical, audio, and video materials), providing classification, prediction, and identification of hidden patterns. On the contrary, generative AI can create (compile) new texts, images, or other data formats based on training samples, which opens up new opportunities to support clinical decision-making and automate workflow and communication processes in healthcare [3–5].

Natural Language Processing (NLP) holds a special place in the modern AI paradigm. It allows to analyze, interpret and generate textual information in a human language. Large Language Models (LLM), specialized AI architectures capable of operating with ultra-large arrays of textual data, have gained development and practical significance. This publication will be

devoted to review of some ethical aspects related to the use of large language models in medicine.

There has been a significant increase in research on the use of LLMs in medical field over the last few years [6,7]. Ethical aspects occupy a central position in discussion about safe and effective implementation of these technologies in clinical practice [8, 9]. Systematic research reveals both the potential advantages of LLMs in medical data analysis, information support, and decision support, significant ethical challenges related to algorithmic bias, lack of transparency, and risks of privacy violations. The ability of LLMs to generate highly persuasive but potentially inaccurate content, which requires human control and development of strict ethical guidelines, is of particular concern.

MATERIALS AND METHODS

While preparing this review publication, an integrated approach was applied to search, analysis and selection of relevant information, including using LLMs. Information search was carried out in domestic and international bibliographic databases: eLibrary, Scopus and PubMed, specialized platforms for searching scientific publications and analytical tools such as Consensus, Semantic Scholar and Elicit, which use LLMs in their algorithms. The search strategy that ensured complete and relevant coverage of the topic under study included key terms and their English-language equivalents such as large language models, medicine, healthcare, ethics, bioethics, risks, bias, reliability, and others. To include sources, a full-text version published in Russian or English from 2015–2025 was required. Relevance of the selected publications according to the abstract was assessed using the following parameters: relevance to the topic of large language models in medicine and healthcare, analysis of ethical aspects, description of implementation risks, novelty and scientific significance of the work. Articles that did not meet the stated criteria, as well as duplicate sources, were excluded. Google's NotebookLM and Perplexity LLM tools were used to systematize, extract data, and summarize selected publications. The resulting prepared materials were checked by a team of authors to ensure accuracy and correctness. The draft was prepared and grammatical proofreading was performed using OpenAI ChatGPT-4.1 and Google Gemini 2.5.

LLM ARCHITECTURE, DEVELOPMENT AND TRAINING

Improvement of computing power, available resources, and advanced algorithms has significantly promoted LLM development, facilitating their integration into various fields of human activity, including clinical practice [1, 4, 5]. LLMs can be used in three main areas such as clinical decision support, automation of medical documentation and reporting, as well as medical education and doctor-patient communication. LLMs have advantages of processing unstructured data [3, 5]. However, effectiveness varies depending on the specific model and approach to training.

To ensure a better understanding of the nature of LLM-related ethical issues, it is necessary to get an idea about the internal structure of the models that shape their functioning specifics. Modern LLMs represent the result of a long-term evolution of architectural approaches in natural language processing. Although transformers have now become a dominant architecture, historically they have gone through several key stages and architectures [10–13]:

1. Early NLP systems were based on manual coding of linguistic rules (for example, the ELIZA system, 1966). Statistical language models (SLM) have been used to predict words based on frequency patterns (for example, IBM Model, 1990). However, they haven't been widely used in practice.
2. Recurrent neural networks (RNNs) include a class of artificial neural nets designed to process sequential information. They can memorize the preceding elements of a sequence. Thus, they can effectively analyze time series, texts, and biomedical signals, operating however with a limited amount of context [11]. Advanced variants with long short-term memory (LSTM) analyze consistent clinical parameters (for example, heart rate, blood pressure, laboratory parameters) and identify patterns that predict complications. LSTMs are used to analyze ECG, EEG, pulse oximetry data, and other time signals [13].
3. Word2Vec implements principles of distributive semantics through vector representations of words (Skip-gram and CBOW algorithms (2013)). In the working process, the text is seen as a sequence of tokens (usually individual words or sub-word units), which are considered as minimal semantic units. For each token, Word2Vec creates an embedding: it maps the token into a multidimensional vector space where words that are similar in meaning have similar vector representations. These embeddings are used to analyze semantic and syntactic relationships in a text.
4. Convolutional neural networks (CNNs) are a class of deep neural networks that initially aimed at processing data with a spatial structure (images, 3D scans, spectrograms). Although CNNs are traditionally associated with image analysis, their architectural principles of extracting local features using convolutional layers served as a prototype for attention mechanisms in transformers, becoming a link between processing local patterns and global context [11, 13].
5. Transformers: a revolutionary architecture based on the mechanism of attention. By using multi-layer encoders/decoders, the model analyzes sequences of tokens, weighing the importance of each token in a sequence. The most well-known models of this class (for example, Generative Pre-trained Transformer, GPT), pre-trained on large-scale text data corpora, became widely available and gained a dominant position [14].
6. Retrieval-Augmented Generation (RAG): an approach aimed at overcoming the fundamental limitations of traditional LLMs, such as generation of factually incorrect information ("hallucinations"), obsolescence of model knowledge and lack of references to verified sources. RAG integrates LLMs with external knowledge bases such as PubMed, UpToDate, clinical recommendation databases, and other reputable resources.
7. BERT (Bidirectional Encoder Representations from Transformers) is an architecture based on bidirectional transformers, which provides a deep understanding of semantics and syntax of the text by taking into account the context to the left and right of the token. BERT and its derivatives are widely used to extract information from electronic medical records, automatically classify medical texts, and get access to biomedical databases and clinical decision support systems.
8. Hybrid models: to solve multimodal problems, systems are being developed that combine transformer attention mechanisms with convolutional or recurrent layers, which allows processing heterogeneous data from text-based

medical records to visualizations (CT, MRI) and time series (ECG, monitoring indicators) [13].

9. Neuro-symbolic systems integrate machine learning methods (neural networks) with symbolic methods of knowledge representation and reasoning (formal logic, expert rules, ontologies). Such systems do not only analyze unstructured data, but also use formal knowledge to improve interpretability, accuracy, and reliability of conclusions. They are used to solve tasks with high requirements for explainability of solutions, for example, when it is necessary to test hypotheses generated by LLMs for compliance with clinical recommendations [15].
10. Reasoning models are designed to solve problems that require complex logical, spatial, or ethical conclusions, optimized to simulate complex cognitive and logical processes typical of medical expertise. Unlike traditional LLMs, which focus primarily on generation of texts and identification of patterns, reasoning models build chains of logical conclusions, integrate diverse sources of knowledge, and explain their decisions at the level similar to the clinical thinking of a professional [16].

A similar path of evolution of technology from basic math algorithms through closed neural network models of "black boxes" gives rise to modern explainable models [16].

MODEL TRAINING

Evolution from rigid linguistic rules and statistical models to modern transformers and hybrid multimodal architectures has significantly expanded the range of LLM application in clinical practice. However, quality and reliability of LLMs directly depend on methods of their training as well as characteristics and quality of the starting training material. In clinical context, it is the initial data that determine the boundaries of the model applicability, level of reliability, interpretability of results, and safety of implementing LLMs in medical processes [17, 18].

Pre-training: the initial stage where the model learns patterns from massive unstructured bodies of general texts. The goal is to form universal language concepts and basic skills for text understanding and generating. Functioning of widely available general-purpose GPT models (YandexGPT, GigaChat, ChatGPT, Gemini, DeepSeek, Grok, Cloud, and others) that can generate different texts, including medical ones, which are however often of a general and superficial nature only is commonly determined at this stage of training. Such models can most likely make mistakes while processing queries concerning complex clinical cases. To avoid potential harm and legal claims, developers equip systems with modules that block responses to medical inquiries, and such an LLM must formulate a disclaimer when responding by recommending you to contact a qualified doctor.

Fine-tuning: additional model training based on specialized clinical data in order to adapt to specific tasks such as generating medical reports, supporting the diagnostic process, analyzing clinical dialogues, processing medical images, etc. Customizable datasets marked up by experts that reflected real clinical scenarios are the most effective. Models that went through such a customization (for example, BioGPT, BioMedLM, PubMedBERT, ClinicalBERT) are commonly used by medical professionals and are less known to the general public [17].

Reinforcement Learning based on Human Feedback (RLHF): a method in which a model corrects its behavior assessing quality and accuracy of the generated responses provided by experts. This minimizes the risk of generating dangerous or incorrect medical recommendations and reducing the likelihood of "hallucinations." Models trained with RLHF (for example, GatorTron, Med-PaLM, MetaMedLLM) are used mainly through integrations that provide access to the context in the form of personalized medical records, electronic health records, integrated and telemedicine solutions. RLHF is approved as the standard for medical LLM training. Research shows that LLMs that used RLHF were superior in quality and completeness of medical consultations compared to both models pre-configured without the RLHF and with pre-trained LLMs. RLHF is an obligatory stage for creation of modern medical language models, ensuring their compliance with requirements of clinical practice, safety and ethics [16].

Quality criteria of the starting training material:

- Relevance and reliability. It is critically important to use only up-to-date and verified data in medical LLMs. Use of outdated or unverified sources can lead to distribution of erroneous recommendations and create risks for the health of patients.
- Representativeness and diversity. To ensure fairness and universality of the model, the training material should cover a wide range of clinical scenarios, demographic groups, linguistic and cultural characteristics. Insufficient representation leads to systematic errors and bias, especially in relation to small or vulnerable groups of patients.
- Markup quality and expert validation. Errors in data annotation, incomplete or incorrect instructions lead to decreased accuracy and interpretability of the results. An effective approach is a combined markup method, in which experts form the core of the dataset, and AI algorithms complement it with variable examples, combining scalability and high-quality annotations.

While performing diagnostics, interpretation of medical images, and clinical communication, models trained on specialized, expertly labeled data demonstrate significantly higher accuracy and stability of results compared to those trained on general or synthetic data sets. [1, 2, 7, 18].

PROBLEMS AND CHALLENGES OF LLM IMPLEMENTATION IN MEDICINE

LLM implementation is accompanied by numerous ethical issues that require a systematic approach to their solution. A comprehensive analysis of LLM-associated ethical challenges has revealed both long-discussed issues such as potential copyright infringement, systematic bias, and data privacy, as well as new dilemmas, including verity of the information generated and its compliance with social norms [1, 8, 9, 18].

COPYRIGHT

As per the classical doctrine of copyright, an author, a person who has a creative idea and implements it in an objective form, can be a natural person only. Emergence of increasingly autonomous AI models capable of generating texts, scientific hypotheses, and diagnostic conclusions raises the question of copyright proprietor [19–23].

In most national legal systems, including the CIS countries, the EU and the USA, copyright does not recognize AI as an independent author (subject). It happens because a creative act needs the presence of will, consciousness and subjective choice, which modern AI does not possess. Article 1228 of the Civil Code of the Russian Federation clearly defines that an author of the work is the citizen (natural person) by whose

creative labor such work of literature, science or art has been created. AI does not have legal capacity and cannot carry out creative activities in the legal sense.

However, the growing volume of medical texts generated by LLMs requires a revision of established approaches. The medical field places special demands on quality, reliability and legal purity of information. Health and life of patients, as well as the professional reputation of medical professionals and researchers, are at stake here unlike artistic or journalistic activities [18]. Use of LLMs for automated creation of medical texts, protocols, data analyses, and even scientific articles generates a number of specific risks:

1. Sources are not obvious: training LLMs require vast amounts of text data, often without a clear distinction between open and copyrighted materials. This hinders identification of sources of borrowings and may lead to an unintended violation of the rights of third parties [20].
2. The problem of plagiarism: automatic text generation can lead to derivative works or texts that partially match the original sources, which poses the threat of accusations of plagiarism from copyright holders.
3. Difficulties with attribution: in case of joint human and AI creativity, it is necessary to determine the contribution of each participant and the order of distribution of copyrights.

There are three main approaches to determation of authorship when creating objects with AI participation [23]:

The author develops AI. It is assumed that all rights to the results created using AI belong to the person or organization that developed the corresponding model. The developer invests significant intellectual efforts and creative potential in the AI system, including development of algorithms, architecture and preparation of data for training. It requires significant financial, time and human resources from the developer [23]. Recognition of copyright by the developer can serve as an incentive for further investments and innovations in this area. This option provides a simpler and more predictable mechanism for determining the copyright holder compared to others. However, this approach is justified only if the user does not make a significant creative contribution, but only presses a button to generate a random piece without further creative intervention.

The author uses AI. In this case, the author is the person who directly manages AI and generates requests. The user chooses from the suggested options, corrects and directs AI to achieve the desired result. A detailed and creative query can lead to a unique piece, while a general or standard query is likely to produce a more typical result. AI acts as an improved tool that allows you to implement the user's creative intent by guiding the process. This model is most often used in medical and legal practice provided that the user (doctor, researcher) is engaged in active participation [22–24].

The author is AI (the concept of "electronic personality"). According to the resolution of the European Parliament with recommendations on civil law rules on robotics, the possibility of recognizing AI as an independent subject of copyright is being discussed [23,24]. Modern generative systems show an increasing degree of autonomy in the process of creating works. Contribution of AI can go beyond a simple instrumental use, and the system is able to generate unexpected and original results that were not directly established by the developer or controlled by a human. However, in practice, this approach has not been recognized, since AI has neither legal personality nor ability to exercise rights and obligations independently. International practice shows that in the vast majority of cases, courts and intellectual property offices refuse to recognize authorship of AI [22].

Thus, we believe that contribution of the participants to creation of any work (literary text, scientific text, and medical records generated by LLMs) is multilevel. When contribution of a user and AI (as a result of developer's work and data) is inseparable, it is necessary to apply the concept of joint authorship, providing compensation to copyright holders depending on their contribution to making content. Depending on the chosen tariff, AI users acquire AI as a service, strengthening their copyright positions.

At the same time, a number of countries are discussing options for introducing special protection regimes for works created with minimal human involvement, for example, a shortened copyright term [21], while providing remuneration to those authors whose works were used to teach AI.

Apart from the legal aspects, the use of LLMs in medicine raises a number of scientific dilemmas/ Reducing the role of human creativity is one of them. Exponential growth in the amount of content generated by AI can devaluate human input and decrease motivation for independent scientific research. Automatic generation of medical texts without proper expert validation can result in distribution of unreliable or even dangerous information.

The modern legal system is not yet ready to fully take into account specifics of AI-generated objects, which requires new approaches to determining authorship, protectability and distribution of rights to the results of intellectual activity.

Taking into account the problems outlined, the following directions of development are proposed:

- Introduction of special protection regimes for works created using AI, for example, a shortened term of rights.
- Mandatory disclosure of AI involvement degree in publication of medical articles, development of clinical protocols and other scientific materials.
- Development of international standards on attribution and identification of sources when using LLMs.
- Creation of more advanced systems for tracking borrowings and checking for plagiarism based on tokenized information.
- Accrual of remuneration to developers and authors of materials on the basis of which models are trained, including through paid subscription systems.

BIAS, HALLUCINATIONS, AND EXPLICABLE AI

Despite significant progress in reducing the frequency of factual errors ("hallucinations") in modern LLMs, especially in highly specialized systems configured using RLHF (with relevance of responses above 95%), a new serious challenge is systematic bias, which leads to errors in medical recommendations, discrimination against vulnerable groups of patients, and distortion of medical knowledge, causing a decreased confidence in AI in healthcare [24,25].

Systematic bias is a persistent distortion of the results of a model due to specific data, architecture, or learning processes, leading to a distorted or inaccurate representation of certain groups, phenomena, or concepts, as well as distorted interpretation of clinical data. These failures are not accidental, they constitute a consequence of the internal algorithm logic. Algorithmic systems cannot only reproduce but also amplify existing biases, creating a potentially dangerous cycle of increased discrimination [26].

LLMs are trained on text corpora that may contain historical, social, and cultural biases, as well as an unbalanced medical knowledge. Errors or subjectivity in marking up medical data can consolidate bias at the stage of preparing datasets.

The features of transformers, attention mechanisms, and ways of processing context can both enhance and weaken bias. As it has already been mentioned, GPT is an autogregressive transformer model trained to predict the next token based on statistical patterns in the training data. It tends to reproduce the most common patterns, reinforcing existing biases and medical stereotypes, which may manifest itself in disproportionate attention to certain aspects of information correlating with demographic characteristics, or in incorrect interpretation of rare or ambiguous cases [25]. GPT has no built-in fact-checking or compliance mechanisms for clinical standards. Increasing the size of the model does not always guarantee less biases; some of its forms may even get intensified [14].

Although reasoning models include logical inference mechanisms (for example, Chain-of-Thought, CoT), they can still reproduce biased reasoning patterns if they were present in the training data, moreover, it is more difficult to detect bias in reasoning chains, because the confirmation bias effect is possible. A critical problem is that the explanations (rationalizations) generated can mask the true (possibly biased) reasons for the model's prediction, especially when the answers are incorrect. The risk reduction approach is to use an expression of uncertainty, where the model indicates the degree of confidence in its response, allowing clinicians to take this into account during interpretation. When models explicitly express their uncertainty, their forecasts become less categorical and less prone to systematic errors [25]. Uncertainty representations can be used as an additional filter to identify cases in which the model is potentially biased or uncertain as it allows either to postpone a decision or involve an expert;

Integration with external knowledge bases in RAG models potentially reduces bias through access to relevant and evidence-based facts. However, RAG models may incorrectly aggregate controversial information from sources or reproduce bias if it is contained in external databases. It is difficult to ensure reproducibility of solutions, because the model may refer to different sources even with identical queries, which makes it difficult to audit and correct bias.

In general, all LLMs are algorithmically inclined to generate the most likely (frequent) responses, ignoring rare but clinically significant cases. When a LLM is used without expert validation, it can lead to perpetuating and spreading bias [14].

Research shows that large language models exhibit significant differences between their "revealed beliefs" and "stated answers," indicating the presence of multiple biases and distortions in the representations they form [26].

Another problem is the dissonance between the probabilistic nature of algorithmic conclusions and their subjective perception by patients (and in some cases by doctors) as deterministic predictions [27].

Research in risk communication confirms that effectiveness of transmitting medical information significantly depends on the way the data is presented to the patient [27]. Categorical formulations of prognostic conclusions induce pronounced psychological reactions even in a low statistical probability of the predicted outcome. Optimistic formulations create the illusion of controllability, forcing patients to underestimate the objective risks and even discontinue therapy prematurely.

Automation bias is the tendency to perceive algorithmic inferences as more objective than human judgments. Digital interfaces make us trust sources subconsciously. Excessive trust in algorithmic advisors is a complex phenomenon of emergence of new forms of dependence. Many users tend to attribute the properties of "superhuman intelligence" to AI systems, ignoring limitations of the training data and architectural features of the models. Experimental data show that 68% of respondents are ready to follow the advice of AI, even though their attending physician has a different opinion [27]. Clinical manifestations of algorithmic dependence include compulsive verification of predictions through mobile applications, anxiety-phobic reactions when the service is temporarily unavailable, and refusal to analyze symptoms independently in favor of automated diagnoses.

Development of methodologies for quantifying bias and degree of reliability of responses in medical LLMs is an important area of further research [28, 29].

Despite the unprecedented potential of LLMs in medicine, their widespread adoption is inhibited by the lack of transparency of decision-making mechanisms for most users, which reduces the trust of medical professionals and patients. Many large language models, such as GPT-4, are complex neural network architectures with billions of parameters, with its internal functioning often being incomprehensible to many users (a "black box") [14].

Explicable Artificial Intelligence (XAI) is a research area focused on development of methodologies and technologies that make the decision-making process of AI systems understandable to humans, enable verification of results and help to overcome distrust in AI technologies [30].

Creating models with initially high degree of interpretability are basic solutions (for example, linear models and decision trees that allow you to explicitly trace the relationship between the input data (the contribution of each feature) and output results). However, these models may have inferior predictive accuracy for some tasks as compared to more complex architectures [16].

Generating intermediate stages of reasoning before giving a final Chain-of-Thought (CoT) response increases not only accuracy, but also explainability, allowing to trace the logical chain of the model. Explanations can be adapted for different groups (doctors, patients, regulators).

As mentioned earlier, it becomes mandatory to apply the RAG methodology, provide models with access to relevant scientific literature, clinical recommendations and other verifiable sources, and increase the accuracy, reliability and transparency of the information generated. The Medical Information Retrieval-Augmented Generation Evaluation (MIRAGE), the first benchmark that includes 7,663 questions from five medical datasets for question-and-answer systems, can serve as an example of an assessment. Studies with MIRAGE have demonstrated that the use of MedRAG, compared with the chain-of-reasoning hint method, improves accuracy of responses from various LLMs by up to 18% [31].

As of May 2025, the MedAgentsBench benchmark includes 1,453 structured clinical cases covering 13 organ systems and 10 medical specialties. According to the comparison results, DeepSeek R1 and OpenAI-o3 reasoning models are the leaders in March 2025. They provide not only high accuracy, but also an optimal ratio between performance, cost of calculations and output time, which is especially important for practical implementation in medical information systems with accuracy in simple diagnostic tasks of 89% for OpenAI-o3 and 93% for DeepSeek R1. However, in complex scenarios requiring multi-stage treatment planning, the indicator decreased to 67% for OpenAI-o3 and 73% for DeepSeek R1 [32].

The problem of lack of standardized metrics and protocols for evaluating the quality of explanations is urgent. Existing XAI methods generate explanations of various formats and content. Currently, there is no consensus on what properties a "good" explanation should have and how these properties can be objectively measured [18,32].

CONFIDENTIALITY AND PROTECTION OF PERSONAL DATA

Use of real clinical data for LLM training and application requires strict adherence to patient anonymization and confidentiality standards, which imposes additional requirements on preparation of training samples [33, 34].

Effectiveness of digital medical technologies directly depends on trust of patients. Violation of confidentiality undermines trust in healthcare system as a whole and can lead to refusal of patients to provide complete and reliable information, which will negatively affect the quality of medical care. Personalized LLMs improve treatment quality by paying attention to individual characteristics, but require processing of ultra-sensitive data (regarding genome, lifestyle, and mental status of the patient) [14].

Medical data can be characterized by a high degree of sensitivity: they contain information about diagnoses, test results, genetic characteristics, medical history, and other information that can identify the patient. They are also subject to strict legal and ethical protection. LLMs are trained on a vast amount of text, including not only open sources, but also specialized medical databases. Even formal depersonalization can be followed by a risk of restoring the patient's identity based on indirect signs, which is especially important for rare diseases or unique combinations of clinical signs.

Order No. 139n of the Ministry of Health of the Russian Federation dated March 20, 2025 "On Approval of the Procedure for Depersonalizing Information about persons who receive medical care, as well as about persons for whom medical expertises, medical examinations and medical certifications are conducted", that has been put in force since September 1, 2025 and that replaced Order No. 341n dated June 14, 2018, prescribes depersonalization of all information that allows direct or indirect identification of the patient's identity, including full name, date of birth, address, contact information, individual document numbers and other identifiers. The procedure should ensure that it is impossible to restore the patient's identity without using additional information stored separately and protected in accordance with the legislation of the Russian Federation [35].

However, even when direct identifiers (name, date of birth, address) are deleted, quasi-identifiers (for example, a rare combination of symptoms, unique treatment regimens) are still present in the medical data and can be used to re-identify the patient. The LLM-Anonymizer study demonstrated retention of about 2% of identifying information after processing [36]. Research shows that intruders can restore source texts from vector representations of models with an accuracy of up to 92% using inversion attack methods [37].

According to ethical standards, minimum required amount of data should be used to achieve the goal. However, LLMs, that use huge datasets for training, often process redundant information, which makes it difficult to control information processing and increases the scope of potential leakage.

In most cases, patients consent to processing of their data for specific purposes of diagnosis, treatment, and scientific research. Classical requirements of completeness of information, voluntary nature, and patient competence conflict with the technical complexity of AI. Use of LLMs capable of generating new knowledge and reusing information in unforeseen scenarios goes beyond the standard forms of consent. Patients are often unaware that their data can be used to train complex models that are subsequently used in a wide range of tasks. Most patients do not have specialized knowledge that allows them to evaluate the architecture of neural networks, quality of training data, or limitations of algorithms [18,34].

LLMs are continuously updated. It makes the traditional static provision of information irrelevant already at the stage of signing the consent. Dynamic informed consent is a modern model of interaction between a patient and a medical organization, which involves not a one-time, but continuous, step-by-step informing of the patient and obtaining the patient's consent at each stage of interaction. The patient obtains information not only at the initial stage of treatment, but also with every significant change in AI algorithm, software update, or occurrence of new clinical data that affect decision-making. It is necessary to use interactive digital platforms that allow the patient to receive notifications, clarifications and consent to new stages of interaction in real time [38,39].

In Russia, there is an experimental legal regime for development and implementation of artificial intelligence (AI) in healthcare, automatically implying consent of patients to transfer anonymized medical data for artificial intelligence training [40], after which the medical community needs to determine the forms and methods of working with dynamic consent.

Existing laws (for example, HIPAA in the USA, GDPR in the EU, FZ-152 in the Russian Federation) establish requirements for personal data protection, but do not take into account the specifics of LLM work. The "right to be forgotten" requirement faces the technical difficulty of selective deletion of data in pre-trained models. There are questions about distribution of responsibility for data leakage (developer, medical institution, user) and compliance with the rules of cross-border data transfer.

Comprehensive regulatory measures are needed: staff training on cybersecurity and ethics of working with medical data, introduction of a multi-level system controlling access to source data and model results, regular testing of models for reproducing sensitive information, introduction of algorithms for detecting and filtering personal data at the stage of generating model responses, use of differential privacy methods that allow training LLMs on aggregated data without the risk of restoring individual records. Legislation needs to be updated considering specifics of LLM work, introduction of special requirements for anonymization and audit of models, and industry standards for certification of depersonalization algorithms. Ensuring transparency of data processing processes and informing patients about possible risks is important too.

Technological solutions such as adding Gaussian noise to embeddings reduce the risk of inversion by 60%, but also worsen performance of the models. Federated Learning (FL) and Homomorphic Encryption (HE) form a technological symbiosis that allows processing sensitive medical data without direct exposure [41].

Federated learning implements a decentralized approach where models are trained on local datasets without their transfer to the central server. It can minimize the risks of leaks in cross-border research and combine knowledge from diverse sources (laboratories, hospitals, wearable devices). Experiments with the Flower FL framework demonstrate high accuracy while significantly reducing privacy risks [42].

Homomorphic encryption schemes make it possible to calculate encrypted data without the need to decrypt it first. Homomorphic encryption means that if the source data has been encrypted, then certain mathematical operations can be done with this ciphertext (for example, addition, multiplication), and the result of these operations will also be encrypted. After decryption of the result, the doctor receives the same result that would have been obtained by performing similar operations with original unencrypted data. However, to optimize these calculations, specialized expensive computing equipment is required [43].

The MedSecureAI prototype demonstrates that the FL+HE combination reduces the risk of leaks by 99.2% while increasing the training time by only 2.1 times compared to the basic models [41]. This creates additional technological challenges: creation of specialized processors for medical HE, development of interstate standards for exchange of encrypted models, and integration of post-quantum cryptographic algorithms.

LEGAL LIABILITY OF LLM RESULTS

From a legal point of view, LLMs currently do not have the status of independent legal entities. They are considered exclusively as tools created and used by individuals or legal entities. The legal responsibility for consequences of LLM application lies with developers, software vendors, as well as medical professionals and organizations using these technologies [44].

Developers and suppliers have to ensure that their products comply with established quality and safety standards, and have to inform users about possible limitations and risks.

All medical devices, including large language model software, are subject to mandatory state registration before they are introduced into clinical practice. Depending on the potential harm caused by an error, AI solutions belong to the following risk classes: IIa (medium risk — systems for pre-processing medical documentation, primary screening), IIb (increased risk — systems for automated interpretation of instrumental research results, algorithms for predicting the course of diseases, software for supporting clinical decision-making) or III (high risk — AI systems that make independent clinical decisions, form diagnostic and therapeutic recommendations, and are applied autonomously in implantable medical devices), since their errors can lead to significant consequences for the patient's life and health. Registration requires conducting a clinical assessment, confirming the quality of algorithms, ensuring transparency and reproducibility of results, as well as implementing risk management mechanisms and continuous monitoring of functioning[45]. Roszdravnadzor monitors and may suspend the use of compromised solutions to take corrective action (as it was done in 2023–2024 with Botkin.AI).

Developers and operating organizations should pay special attention to information security issues. Information systems that process personal data of patients are becoming a priority target for intruders. Modern cyber security threats, including unauthorized access, attacking integrity and confidentiality of data, as well as manipulation with model conclusions, can lead to serious consequences that pose threats not only to health, but also to lives of patients [45, 46]. These medical information systems are subject to Federal Law No. 187-FZ dated July 26, 2017 "On Security of Critical Information Infrastructure of the Russian Federation".

Medical professionals, in turn, are professionally responsible for making clinical decisions, even if they rely on recommendations formulated by the LLM. The doctor must critically evaluate the information received and cannot completely delegate decision-making to artificial intelligence [18].

In case of negative consequences related to errors or unreliable LLM recommendations, responsibility can be distributed among various participants of the process, depending on the nature and source of the error. If we are talking about a software defect, responsibility is usually allocated to the developer. The medical professional or organization is claimed responsible if an error occurred due to incorrect use of technology or because a doctor ignored professional standards and clinical recommendations.

CONCLUSION

Thus, introduction of large language models in healthcare requires an integrated approach combining further technological improvement of models, development and implementation of strict ethical standards, adaptation of the regulatory framework, use of advanced information security techniques and constant critical supervision by the expert medical community.

Improvement of algorithms and architectures is one of the key areas. It is necessary to select modern models that combine the possibilities of reasoning, search and explanation. Transition from predictive "black box" models to interpreted systems that can substantiate their conclusions will increase trust of medical professionals and patients in these technologies. Development of neuro-symbolic methods that integrate machine learning with symbolic representations of knowledge and logical reasoning is an important step. It helps not only analyze unstructured data, but also use formal knowledge to improve interpretability, accuracy, and reliability of conclusions.

Quality and relevance of the training data are equally important. LLMs should not only be pre-trained on massive bodies of texts, but also pre-tuned using highly specialized pre-marked clinical data with participation of medical experts. Expert Feedback Reinforcement Learning (RLHF) should become the standard for medical language models, confirming their compliance with requirements of clinical practice, safety and ethics. This will ensure not only the relevance of general answers, but also their personification and clinical evidence.

Adjustment of regulatory framework to technological advances is a prerequisite for successful implementation of LLMs in healthcare. Legal experts need to consider specifics of increasing AI integration into all fields of activity and develop new approaches to authorship identification, protection and distribution of rights to intellectual property results created with LLM participation.

Ensuring confidentiality and protection of personal data is a prerequisite. It is important to strictly adhere to the standards of patient anonymization and confidentiality when using real clinical data for LLM training and application. The minimum required amount of data should be used to achieve the goal set and implement technological solutions such as federated learning and homomorphic encryption that allow to process sensitive medical data without direct exposure. It is also important to develop interactive digital platforms that provide the patient with real-time notifications, clarifications and consent to new stages of interaction (dynamic consent form).

Exclusion of unreliable answers, step-by-step fact-checking and cross-checking are necessary to combat "hallucinations" and bias. It is necessary to develop methodologies to quantify the degree of reliability of responses in medical LLMs, allowing clinicians to take this into account when interpreting the results. It is important to pay attention to cultural and linguistic characteristics of different groups of patients and develop models that take these differences into account.

To develop an objective and trusting attitude towards the applied AI technologies, it is necessary to ensure transparency and explainability of LLM functioning. To do this, it is necessary to develop standardized metrics and protocols assessing quality and use of XAI methods to trace the logical chain of the model and adapt explanations for different audiences. It is also important to take into account psychological aspects of LLM-provided information perception and avoid categorical formulations that can induce pronounced psychological reactions.

Only when these conditions are met, healthcare level can be significantly increased owing to the use of large language models, while protecting the rights and interests of patients and medical professionals.

References

1. Meng X, Yan X, Zhang K, et al. The application of large language models in medicine: A scoping review. iScience. 2024; 27(5): 109713.
2. Omiye JA, Gui H, Rezaei SJ, Zou J, Daneshjou R. Large Language Models in Medicine: The Potentials and Pitfalls: A Narrative Review. Ann Intern Med. 2024; 177(2): 210–220. DOI: 10.7326/M23-2772.
3. Chen Y, Esmaeilzadeh P. Generative AI in Medical Practice: In-Depth Exploration of Privacy and Security Challenges. J Med Internet Res. 2024; 26: e53008. DOI: 10.2196/53008.
4. Yim D, Khuntia J, Parameswaran V, Meyers A. Preliminary Evidence of the Use of Generative AI in Health Care Clinical Services: Systematic Narrative Review. JMIR Med Inform. 2024; 12: e52073. DOI: 10.2196/52073.
5. Nógrádi B, Polgár TF, Meszlényi V, et al. ChatGPT M. D. Is there any room for generative AI in neurology? PLoS One. 2024; 19(10): e0310028. DOI: 10.1371/journal.pone.0310028.
6. Wang C, Li M, He J, Wang Z, Darzi E, Chen Z, et al. A Survey for Large Language Models in Biomedicine. ArXiv. 2024; abs/2409.00133.
7. Moglia A, Georgiou K, Cerveri P, et al. Large language models in healthcare: from a systematic review on medical examinations to a comparative analysis on fundamentals of robotic surgery online test. Artif Intell Rev. 2024; 57: 231. DOI: 10.1007/s10462-024-10849-5.
8. Ong JCL, Chang SY, William W, et al. Ethical and regulatory challenges of large language models in medicine. Lancet Digit Health. 2024; 6(6): e428-e432. DOI: 10.1016/S2589-7500(24)00061-X.
9. Zhui LL, Fenghe L, Xuehu W, Qining F, Wei R. Ethical Considerations and Fundamental Principles of Large Language Models in Medical Education: Viewpoint. J Med Internet Res. 2024; 26: e60083. DOI: 10.2196/60083
10. Wei Y, Zhou J, Wang Y, et al. A Review of Algorithm & Hardware Design for AI-Based Biomedical Applications. IEEE Trans Biomed Circuits Syst. 2020; 14(2): 145–163. DOI: 10.1109/TBCAS.2020.2974154
11. Pikalov Ya S. Obzor arkhitektur sistem intellektual'noy obrabotki yestestvenno-yazykovykh tekstov. Problemy iskusstvennogo intellekta. 2020; 4(19). Russian.
12. Jin L, Feng S, Xin Z, Chai Y. Evolution and advancements in deep learning models for Natural Language Processing. Applied and Computational Engineering. 2024.
13. Li K, Ao B, Wu X, Wen Q, Ul Haq E, Yin J. Parkinson's disease detection and classification using EEG based on deep CNN-LSTM model. Biotechnol Genet Eng Rev. 2024; 40(3): 2577–2596. DOI: 10.1080/02648725.2023.2200333
14. Haltaufderheide J, Ranisch R. The ethics of ChatGPT in medicine and healthcare: a systematic review on Large Language Models (LLMs). NPJ Digit Med. 2024; 7(1): 183. Published 2024 Jul 8. DOI: 10.1038/s41746-024-01157-x.
15. Garcez AA, et al. Neural-symbolic computing: An effective methodology for principled integration of machine learning and reasoning. arXiv preprint arXiv:1905.06088. 2019.
16. Li Z, et al. From system 1 to system 2: A survey of reasoning large language models. arXiv preprint arXiv: 2502.17419. 2025.
17. Cascella M, Semeraro F, Montomoli J, Bellini V, Piazza O, Bignami E. The Breakthrough of Large Language Models Release for Medical Applications: 1-Year Timeline and Perspectives. J Med Syst. 2024; 48(1): 22. DOI: 10.1007/s10916-024-02045-3.
18. Andreychenko AYe., Gusev AV. Perspektivy primeneniya bol'shikh yazykovykh modeley v zdravookhranenii. Natsional'noye zdravookhraneniye. 2023; 4 (4): 48–55. DOI: 10.47093/2713-069X.2023.4.4.48-55. Russian.
19. Moffatt B, Hall A. Is AI my co-author? The ethics of using artificial intelligence in scientific publishing. Account Res. 2024. DOI: 10.1080/08989621.2024.2386285.
20. Lee JY. Can an artificial intelligence chatbot be the author of a scholarly article? J Educ Eval Health Prof. 2023; 20: 6. DOI: 10.3352/jeehp.2023.20.6.
21. Johnson A. Generative AI, UK Copyright and Open Licences: considerations for UK HEI copyright advice services. F1000Res. 2024; 13: 134. DOI: 10.12688/f1000research.143131.1.
22. Shaydurov AS. Avtorskoye pravo na proizvedeniya, sozdannyye iskusstvennym intellektom v RF: problemy i perspektivy. V sbornike: Tsifrovyye tekhnologii v nauchnom razvitii: novyye kontseptual'nyye podkhody: Sbornik statey po itogam Mezhdunarodnoy nauchno-prakticheskoy konferentsii; 30 aprelya 2023 g.; Samara. Sterlitamak: Obshchestvo s ogranichennoy otvetstvennost'yu «Agentstvo mezhdunarodnykh issledovaniy». 2023; 88–92. EDN MDZVBE. Russian.
23. Kishkembayev M. Aktual'nyye problemy zashchity avtorskikh prav na ob"yekt, sozdannyy iskusstvennym intellektom. Vestnik Torajgyrov Universiteta. Ûridičeskaâ Seriâ. Mart 2024; 75–86. DOI: 10.48081/zusz1934. Russian.
24. Avila Negri SMC. Robot as Legal Person: Electronic Personhood in Robotics and Artificial Intelligence. Front Robot AI. 2021;8:789327. Published 2021 Dec 23. DOI: 10.3389/frobt.2021.789327.
25. Schmidgall S, Harris C, Essien I, et al. Evaluation and mitigation of cognitive biases in medical language models. NPJ Digit Med. 2024; 7(1): 295. DOI: 10.1038/s41746-024-01283-6.
26. Mondal M, et al. Do large language models exhibit cognitive dissonance? studying the difference between revealed beliefs and stated answers. Preprint arXiv. 2406.14986. 2024.
27. Goerlandt F, Li J, Reniers G. The Landscape of Risk Communication Research: A Scientometric Analysis. International Journal of Environmental Research and Public Health. 2020; 17(9): 3255. DOI: 10.3390/ijerph17093255.
28. Zarfati M, Soffer S, Nadkarni GN, Klang E. Generation: Advancing personalized care and research in oncology. Eur J Cancer. 2025; 220: 115341. DOI: 10.1016/j.ejca.2025.115341.
29. Shool S, Adimi S, Saboori Amleshi R, Bitaraf E, Golpira R, Tara M. A systematic review of large language model (LLM) evaluations in clinical medicine. BMC Med Inform Decis Mak. 2025; 25(1): 117. DOI: 10.1186/s12911-025-02954-4.
30. Shevskaya NV. Ob"yasnimyy iskusstvennyy intellekt i metody interpretatsii rezul'tatov. Modelirovaniye, optimizatsiya i informatsionnyye tekhnologii. 2021; 9(2). DOI: 10.26102/2310-6018/2021.33.2.024. EDN VRKUIL. Russian.
31. Xiong G, et al. Benchmarking retrieval-augmented generation for medicine. Findings of the Association for Computational Linguistics ACL. 2024. 2024; 6233–6251.
32. Tang X et al. Medagentsbench: Benchmarking thinking models and agent frameworks for complex medical reasoning. arXiv preprint arXiv:2503.07459. 2025.
33. Yadav N, Pandey S, Gupta A, Dudani P, Gupta S, Rangarajan K. Data Privacy in Healthcare: In the Era of Artificial Intelligence. Indian Dermatol Online J. 2023;14(6): 788–792. DOI: 10.4103/idoj.idoj_543_23.

34. Altynnikov MS, Kuznetsova NO. Osobennosti organizatsii zashchity personal'nykh dannykh v meditsinskom uchrezhdenii. Innovatsii. Nauka. Obrazovaniye. 2021; 36: 1479–1486. EDN ZISIGQ. Russian.
35. Prikaz Minzdrava Rossii ot 20.03.2025 № 139n «Ob utverzhdenii Poryadka obezlichivaniya svedeniy o litsakh, kotorym okazyvayetsya meditsinskaya pomoshch', a takzhe o litsakh, v otnoshenii kotorykh provodyatsya meditsinskiye ekspertizy, meditsinskiye osmotry i meditsinskiye osvidetel'stvovaniya». Ofitsial'nyy internet-portal pravovoy informatsii. 2025. Russian.
36. Wiest IC, et al. Anonymizing medical documents with local, privacy preserving large language models: The LLM-Anonymizer. medRxiv. 2024. C. 2024.06. 11.24308355.
37. Morris JX, et al. Text embeddings reveal (almost) as much as text. arXiv preprint arXiv:2310.06816. 2023.
38. Mascalzoni D, Melotti R, Pattaro C, et al. Ten years of dynamic consent in the CHRIS study: informed consent as a dynamic process. Eur J Hum Genet. 2022;30: 1391–1397 DOI: 10.1038/s41431-022-01160-4. 2022.
39. Harishbhai Tilala M, Kumar Chenchala P, Choppadandi A, et al. Ethical Considerations in the Use of Artificial Intelligence and Machine Learning in Health Care: A Comprehensive Review. Cureus. 2024; 16(6): e62443. Published 2024 Jun 15. DOI: 10.7759/cureus.62443.
40. Postanovleniye Pravitel'stva RF ot 18.07.2023 № 1164 (red. ot 01.02.2025) «Ob ustanovlenii eksperimental'nogo pravovogo rezhima v sfere tsifrovykh innovatsiy i utverzhdenii Programmy eksperimental'nogo pravovogo rezhima v sfere tsifrovykh innovatsiy po napravleniyu meditsinskoy deyatel'nosti, v tom chisle s primeneniyem telemeditsinskikh tekhnologiy i tekhnologiy sbora i obrabotki svedeniy o sostoyanii zdorov'ya i diagnozakh grazhdan». Ofitsial'nyy internet-portal pravovoy informatsii. 2025. Russian.
41. Lessage X, Collier L, Van Ouytsel C-HB, Legay A, Mahmoudi S and Massonet P. Secure federated learning applied to medical imaging with fully homomorphic encryption. 2024 IEEE 3rd International Conference on AI in Cybersecurity (ICAIC). Houston. TX. USA. 2024; 1–12. DOI: 10.1109/ICAIC60265.2024.10433836.
42. Walskaar, I, Tran, MC, & Catak, FO. A Practical Implementation of Medical Privacy-Preserving Federated Learning Using Multi-Key Homomorphic Encryption and Flower Framework. Cryptography. 2023; 7(4): 48. DOI: 10.3390/cryptography7040048.
43. Mohandas R, Veena S, Kirubasri G. Thusnavis Bella Mary I & Udayakumar, R. Federated Learning with Homomorphic Encryption for Ensuring Privacy in Medical Data. Indian Journal of Information Sources and Services. 2024; 14(2): 17–23. DOI: 10.51983/ijiss-2024.14.2.03.
44. Shumway DO, Hartman HJ. Medical malpractice liability in large language model artificial intelligence: legal review and policy recommendations. J Osteopath Med. 2024; 124(7): 287–290. DOI: 10.1515/jom-2023-0229.
45. Tret'yakova Ye P. Ispol'zovaniye iskusstvennogo intellekta v zdravookhranenii: raspredeleniye otvetstvennosti i riskov. Tsifrovoye pravo. 2021; 2(4): 51–60. DOI: 10.38044/2686-9136-2021-2-4-51-60. EDN NGHUTA. Russian.
46. Mazhuga YeYu, Petrova AO, Gordeyev Ya I. Pravovoye regulirovaniye sistem iskusstvennogo intellekta v meditsinskoy sfere. V sbornike: Aktual'nyye problemy sovremennoy Rossii: psikhologiya, pedagogika, ekonomika, upravleniye i pravo. Sbornik nauchnykh trudov mezhdunarodnykh nauchno-prakticheskikh konferentsiy; 07–24 aprelya 2023 g.; Moskva. Moskva: Moskovskiy psikhologo-sotsial'nyy universitet. 2023; 1281–1285. EDN IAGUFA. Russian.

Литература

1. Meng X, Yan X, Zhang K, et al. The application of large language models in medicine: A scoping review. iScience. 2024; 27(5): 109713.
2. Omiye JA, Gui H, Rezaei SJ, Zou J, Daneshjou R. Large Language Models in Medicine: The Potentials and Pitfalls: A Narrative Review. Ann Intern Med. 2024; 177(2): 210–220. DOI: 10.7326/M23-2772.
3. Chen Y, Esmaeilzadeh P. Generative AI in Medical Practice: In-Depth Exploration of Privacy and Security Challenges. J Med Internet Res. 2024; 26: e53008. DOI: 10.2196/53008.
4. Yim D, Khuntia J, Parameswaran V, Meyers A. Preliminary Evidence of the Use of Generative AI in Health Care Clinical Services: Systematic Narrative Review. JMIR Med Inform. 2024; 12: e52073. DOI: 10.2196/52073.
5. Nógrádi B, Polgár TF, Meszlényi V, et al. ChatGPT M. D. Is there any room for generative AI in neurology? PLoS One. 2024; 19(10): e0310028. DOI: 10.1371/journal.pone.0310028.
6. Wang C, Li M, He J, Wang Z, Darzi E, Chen Z, et al. A Survey for Large Language Models in Biomedicine. ArXiv. 2024; abs/2409.00133.
7. Moglia A, Georgiou K, Cerveri P, et al. Large language models in healthcare: from a systematic review on medical examinations to a comparative analysis on fundamentals of robotic surgery online test. Artif Intell Rev. 2024; 57: 231. DOI: 10.1007/s10462-024-10849-5.
8. Ong JCL, Chang SY, William W, et al. Ethical and regulatory challenges of large language models in medicine. Lancet Digit Health. 2024; 6(6): e428-e432. DOI: 10.1016/S2589-7500(24)00061-X.
9. Zhui LL, Fenghe L, Xuehu W, Qining F, Wei R. Ethical Considerations and Fundamental Principles of Large Language Models in Medical Education: Viewpoint. J Med Internet Res. 2024; 26: e60083. DOI: 10.2196/60083
10. Wei Y, Zhou J, Wang Y, et al. A Review of Algorithm & Hardware Design for AI-Based Biomedical Applications. IEEE Trans Biomed Circuits Syst. 2020; 14(2): 145–163. DOI: 10.1109/TBCAS.2020.2974154
11. Пикалев Я. С. Обзор архитектур систем интеллектуальной обработки естественно-языковых текстов. Проблемы искусственного интеллекта. 2020; 4(19).
12. Jin L, Feng S, Xin Z, Chai Y. Evolution and advancements in deep learning models for Natural Language Processing. Applied and Computational Engineering. 2024.
13. Li K, Ao B, Wu X, Wen Q, Ul Haq E, Yin J. Parkinson's disease detection and classification using EEG based on deep CNN-LSTM model. Biotechnol Genet Eng Rev. 2024; 40(3): 2577–2596. DOI: 10.1080/02648725.2023.2200333
14. Haltaufderheide J, Ranisch R. The ethics of ChatGPT in medicine and healthcare: a systematic review on Large Language Models (LLMs). NPJ Digit Med. 2024; 7(1): 183. Published 2024 Jul 8. DOI: 10.1038/s41746-024-01157-x.
15. Garcez AA, et al. Neural-symbolic computing: An effective methodology for principled integration of machine learning and reasoning. arXiv preprint arXiv:1905.06088. 2019.
16. Li Z, et al. From system 1 to system 2: A survey of reasoning large language models. arXiv preprint arXiv: 2502.17419. 2025.
17. Cascella M, Semeraro F, Montomoli J, Bellini V, Piazza O, Bignami E. The Breakthrough of Large Language Models Release for Medical Applications: 1-Year Timeline and Perspectives. J Med Syst. 2024; 48(1): 22. DOI: 10.1007/s10916-024-02045-3.
18. Андрейченко А. Е., Гусев А. В. Перспективы применения больших языковых моделей в здравоохранении. Национальное здравоохранение. 2023; 4 (4): 48–55. DOI: 10.47093/2713-069X.2023.4.4.48-55.
19. Moffatt B, Hall A. Is AI my co-author? The ethics of using artificial intelligence in scientific publishing. Account Res. 2024. DOI: 10.1080/08989621.2024.2386285.
20. Lee JY. Can an artificial intelligence chatbot be the author of a scholarly article? J Educ Eval Health Prof. 2023; 20: 6. DOI: 10.3352/jeehp.2023.20.6.
21. Johnson A. Generative AI, UK Copyright and Open Licences: considerations for UK HEI copyright advice services. F1000Res. 2024; 13: 134. DOI: 10.12688/f1000research.143131.1.

22. Шайдуров А. С. Авторское право на произведения, созданные искусственным интеллектом в РФ: проблемы и перспективы. В сборнике: Цифровые технологии в научном развитии: новые концептуальные подходы: Сборник статей по итогам Международной научно-практической конференции; 30 апреля 2023 г.; Самара. Стерлитамак: Общество с ограниченной ответственностью «Агентство международных исследований». 2023; 88–92. EDN MDZVBE.
23. Кишкембаев М. Актуальные проблемы защиты авторских прав на объект, созданный искусственным интеллектом. Vestnik Torajgyrov Universiteta. Ûridičeskaâ Seriâ. Март 2024; 75–86. DOI: 10.48081/zusz1934.
24. Avila Negri SMC. Robot as Legal Person: Electronic Personhood in Robotics and Artificial Intelligence. Front Robot AI. 2021;8:789327. Published 2021 Dec 23. DOI: 10.3389/frobt.2021.789327.
25. Schmidgall S, Harris C, Essien I, et al. Evaluation and mitigation of cognitive biases in medical language models. NPJ Digit Med. 2024; 7(1): 295. DOI: 10.1038/s41746-024-01283-6.
26. Mondal M, et al. Do large language models exhibit cognitive dissonance? studying the difference between revealed beliefs and stated answers. Preprint arXiv. 2406.14986. 2024.
27. Goerlandt F, Li J, Reniers G. The Landscape of Risk Communication Research: A Scientometric Analysis. International Journal of Environmental Research and Public Health. 2020; 17(9): 3255. DOI: 10.3390/ijerph17093255.
28. Zarfati M, Soffer S, Nadkarni GN, Klang E. Retrieval-Augmented Generation: Advancing personalized care and research in oncology. Eur J Cancer. 2025; 220: 115341. DOI: 10.1016/j.ejca.2025.115341.
29. Shool S, Adimi S, Saboori Amleshi R, Bitaraf E, Golpira R, Tara M. A systematic review of large language model (LLM) evaluations in clinical medicine. BMC Med Inform Decis Mak. 2025; 25(1): 117. DOI: 10.1186/s12911-025-02954-4.
30. Шевская Н. В. Объяснимый искусственный интеллект и методы интерпретации результатов. Моделирование, оптимизация и информационные технологии. 2021; 9(2). DOI: 10.26102/2310-6018/2021.33.2.024. EDN VRKUIL.
31. Xiong G, et al. Benchmarking retrieval-augmented generation for medicine. Findings of the Association for Computational Linguistics ACL. 2024. 2024; 6233–6251.
32. Tang X et al. Medagentsbench: Benchmarking thinking models and agent frameworks for complex medical reasoning. arXiv preprint arXiv:2503.07459. 2025.
33. Yadav N, Pandey S, Gupta A, Dudani P, Gupta S, Rangarajan K. Data Privacy in Healthcare: In the Era of Artificial Intelligence. Indian Dermatol Online J. 2023;14(6): 788–792. DOI: 10.4103/idoj.idoj_543_23.
34. Алтынников М. С., Кузнецова Н. О. Особенности организации защиты персональных данных в медицинском учреждении. Инновации. Наука. Образование. 2021; 36: 1479–1486. EDN ZISIGQ.
35. Приказ Минздрава России от 20.03.2025 № 139н «Об утверждении Порядка обезличивания сведений о лицах, которым оказывается медицинская помощь, а также о лицах, в отношении которых проводятся медицинские экспертизы, медицинские осмотры и медицинские освидетельствования». Официальный интернет-портал правовой информации. 2025.
36. Wiest IC, et al. Anonymizing medical documents with local, privacy preserving large language models: The LLM-Anonymizer. medRxiv. 2024. С. 2024.06. 11.24308355.
37. Morris JX, et al. Text embeddings reveal (almost) as much as text. arXiv preprint arXiv:2310.06816. 2023.
38. Mascalzoni D, Melotti R, Pattaro C, et al. Ten years of dynamic consent in the CHRIS study: informed consent as a dynamic process. Eur J Hum Genet. 2022;30: 1391–1397 DOI: 10.1038/s41431-022-01160-4. 2022.
39. Harishbhai Tilala M, Kumar Chenchala P, Choppadandi A, et al. Ethical Considerations in the Use of Artificial Intelligence and Machine Learning in Health Care: A Comprehensive Review. Cureus. 2024; 16(6): e62443. Published 2024 Jun 15. DOI: 10.7759/cureus.62443.
40. Постановление Правительства РФ от 18.07.2023 № 1164 (ред. от 01.02.2025) «Об установлении экспериментального правового режима в сфере цифровых инноваций и утверждении Программы экспериментального правового режима в сфере цифровых инноваций по направлению медицинской деятельности, в том числе с применением телемедицинских технологий и технологий сбора и обработки сведений о состоянии здоровья и диагнозах граждан». Официальный интернет-портал правовой информации. 2025.
41. Lessage X, Collier L, Van Ouytsel C-HB, Legay A, Mahmoudi S and Massonet P. Secure federated learning applied to medical imaging with fully homomorphic encryption. 2024 IEEE 3rd International Conference on AI in Cybersecurity (ICAIC). Houston. TX. USA. 2024; 1–12. DOI: 10.1109/ICAIC60265.2024.10433836.
42. Walskaar, I, Tran, MC, & Catak, FO. A Practical Implementation of Medical Privacy-Preserving Federated Learning Using Multi-Key Homomorphic Encryption and Flower Framework. Cryptography. 2023; 7(4): 48. DOI: 10.3390/cryptography7040048.
43. Mohandas R, Veena S, Kirubasri G. Thusnavis Bella Mary I & Udayakumar, R. Federated Learning with Homomorphic Encryption for Ensuring Privacy in Medical Data. Indian Journal of Information Sources and Services. 2024; 14(2): 17–23. DOI: 10.51983/ijiss-2024.14.2.03.
44. Shumway DO, Hartman HJ. Medical malpractice liability in large language model artificial intelligence: legal review and policy recommendations. J Osteopath Med. 2024; 124(7): 287–290. DOI: 10.1515/jom-2023-0229.
45. Третьякова Е. П. Использование искусственного интеллекта в здравоохранении: распределение ответственности и рисков. Цифровое право. 2021; 2(4): 51–60. DOI: 10.38044/2686-9136-2021-2-4-51-60. EDN NGHUTA.
46. Мажуга Е. Ю., Петрова А. О., Гордеев Я. И. Правовое регулирование систем искусственного интеллекта в медицинской сфере. В сборнике: Актуальные проблемы современной России: психология, педагогика, экономика, управление и право. Сборник научных трудов международных научно-практических конференций; 07–24 апреля 2023 г.; Москва. Москва: Московский психолого-социальный университет. 2023; 1281–1285. EDN IAGUFA.

ETHICAL AND LEGAL IMPLICATIONS OF GENETIC TESTING IN TRAUMATOLOGY AND ORTHOPEDICS

Savgachev VV ✉

Yaroslavl State Medical University, Yaroslavl, Russia

The field of medicine has traditionally focused on such symptoms of a disease as pain, inflammation, and tissue deformity. However, according to the modern approach, it is necessary to identify the cause of the disease, which is often hidden deep inside the body. Examining genetic polymorphism as the basis for purulent complications after treatment of the lower limb injury was one of the method that could solve the problem. Epidemiological observations confirm that purulent complications after orthopedic surgery are associated with hereditary predisposition factors. This highlights the important role of genetic changes in development and course of this pathology. However, any medical intervention is associated with potential risks, including emotional pain of the patient, violation of personal data confidentiality and misuse of the obtained information. That is why it is important to think in advance about the possible consequences of genetic tests and to find ways how to resolve ethical issues.

Keywords: genetic research, polymorphism, ethics, lower limb injury

Acknowledgements: the author expresses gratitude to Alexander L. Khokhlov, Academician of the Russian Academy of Sciences, Rector of Yaroslavl State Medical University (YSMU), for his help in organizing the laboratory genetic stage of the study.

Compliance with ethical standards: the study was carried out as part of the doctoral dissertation entitled “Treatment of purulent complications and prediction of their outcomes in patients with trauma to the medial lower extremity” approved by the Academic Council of YSMU (extract from protocol No. 11 of the meeting of the Academic Council of YSMU dated 06/26/2024, state registration number: 124071000009-0). The work was approved by the Ethics Committee (extract from the minutes of the meeting of the YSMU Ethics Committee dated June 14, 2024 No. 68).

✉ **Correspondence should be addressed:** Vitaly V. Savgachev
Revolutsionnaya St., 5, Yaroslavl region, Yaroslavl, 150000, Russia; hirurg2288@mail.ru

Received: 30.04.2025 **Accepted:** 12.05.2025 **Published online:** 24.05.2025

DOI: 10.24075/medet.2025.001

ЭТИЧЕСКИЕ И ПРАВОВЫЕ АСПЕКТЫ ПРИ ПРОВЕДЕНИИ ГЕНЕТИЧЕСКОГО ИССЛЕДОВАНИЯ В ТРАВМАТОЛОГИИ И ОРТОПЕДИИ

В. В. Савгачев ✉

Ярославский государственный медицинский университет, Ярославль, Россия

Традиционно медицина уделяла внимание лишь симптомам заболевания — боли, воспалению, деформации тканей. Но современный подход требует выявления причины болезни, часто скрытой глубоко внутри организма. Одним из методов решения этой проблемы стало изучение генетического полиморфизма как основы формирования гнойных осложнений после лечения травмы нижних конечностей. Эпидемиологические наблюдения подтверждают связь возникновения гнойных осложнений после проведенных хирургических вмешательств в ортопедической практике с факторами наследственной предрасположенности. Это подчеркивает важную роль генетических изменений в развитии и течении данной патологии. Однако любые медицинские вмешательства связаны с потенциальными рисками, включая эмоциональные переживания пациента, нарушение конфиденциальности личных данных и неправильное использование полученной информации. Именно поэтому важно заранее учитывать возможные последствия проведения генетических тестов и найти пути разрешения этических вопросов.

Ключевые слова: генетическое исследование, полиморфизм, этика, травма нижних конечностей

Благодарность: автор выражает благодарность за помощь в организации проведения лабораторного генетического этапа исследования академику РАН, ректору Ярославского государственного медицинского университета Александру Леонидовичу Хохлову.

Соблюдение этических стандартов: исследование выполнено в рамках выполнения утвержденной на ученом совете ЯГМУ докторской диссертации на тему «Лечение гнойных осложнений и прогнозирование их исходов у пациентов с травмой среднедистального отдела нижней конечности» (выписка из протокола № 11 заседания ученого совета ЯГМУ от 26.06.2024, номер госрегистрации: 124071000009-0). Работа одобрена этическим комитетом (выписка из протокола заседания этического комитета ЯГМУ от 14 июня 2024 г. № 68).

✉ **Для корреспонденции:** Виталий Владимирович Савгачев
ул. Революционная, д. 5, Ярославская область, г. Ярославль, 150000, Россия; hirurg2288@mail.ru

Статья поступила: 30.04.2025 **Статья принята к печати:** 12.05.2025 **Опубликована онлайн:** 24.05.2025

DOI: 10.24075/medet.2025.001

BACKGROUND

To understand how genetic components can influence occurrence and course of purulent complications in trauma patients, scientists apply comprehensive research strategies, including family history, identification of differences in gene expression and a detailed study of the structural features of the genome [1].

Understanding the individual genetic characteristics of patients helps to create personalized treatment protocols that enhance effectiveness of therapeutic measures [2]. Though the mechanism of complications is still poorly understood, it obviously includes multiple factors and regulatory mechanisms [3]. Individual features of gene functioning have a little impact on the overall risk of disease development in the population. The significance of a specific genetic variation depends on its interaction with external conditions, influence of environmental factors and changes as a result of epigenetic processes characterized by unique physiological reactions in each patient [4].

The study of the impact produced by different genes on formation of a pathogenic cascade leading to purulent complications constitutes a promising trend. Taking into account the key role of immunity in the pathogenesis of purulent complications, genes involved in regulating the immune response are of particular interest. They include IL-17A, which is responsible for synthesis of protein interleukin 17A, and IL-6 that controls cytokine production. Both of these genes are essential in triggering and maintaining inflammation, and are also associated with a variety of diseases, including autoimmune disorders, infections, and cancer [5]. Thus, the significance of genetic research in trauma patients cannot be questioned. However, it is important to consider various ethical and legal implications. Both Russian, and foreign authors highlighted the complexity of resolving ethical issues while dealing with this type of research [6, 7].

ETHICS, MATERIALS AND METHODS OF GENETIC RESEARCH

At the first stage of selecting patients for genetic testing, each potential participant obtained the following information.

The study will be conducted at the laboratory of the Federal State Budgetary Educational Institution of Higher Education Yaroslavl State Medical University of the Ministry of Health of the Russian Federation (Rector, Professor, Academician of the Russian Academy of Sciences Khokhlov AL). A candidate genetic study will be performed. The study is about identification of IL17A and IL6 gene polymorphisms in the genome of patients treated for purulent complications after therapy of lower limb injury using allele-specific polymerase chain reaction (ASPCR) with SNP-EXPRESS-RV, a real-time fluorescent product detection system. This study is being carried out as part of a doctoral dissertation approved by the Academic Council of YSMU and entitled as ''Treatment of purulent complications and prediction of their outcomes in patients with medial lower extremity injuries" (the work was approved by the ethics committee (extract from the minutes of the meeting of the YSMU Ethics Committee dated June 14, 2024 No. 68)).

Inclusion criteria:

1) patients with a history of acute injuries of the musculoskeletal system at the level of the knee joint, shin, ankle and foot (fractures, dislocations, ligament tears);
2) patients who developed purulent and inflammatory complications (osteomyelitis, purulent arthritis, phlegmon, abscess) in the course of therapy and within a year after treatment;
3) the patients are 18–75 years old;
4) informed consent of the patient to participate in the study.

Exclusion criteria:

1) patients with chronic diseases that may influence the course of purulent complications (oncological diseases, HIV infection);
2) patients receiving immunosuppressive therapy;
3) patients with a recurrence rate of purulent and inflammatory complications for less than a year;
4) patients with a known history of genetic diseases associated with impaired immunity;
5) patients who refused to participate in the study.

The materials for the analysis were as follows.

Samples: 2000 µl sample of whole venous blood collected for testing into a disposable Lab-Vac vacuum tube with 200 µl of ethylenediaminetetraacetate anticoagulant solution.

Reagents: a set of reagents for testing single nucleotide polymorphism of the G-197A polymorphism of IL17A with ASPCR Mutation of IL-17A SNP-express-RV undiscovered-100 (Litekh, Russia); a set of reagents for testing single nucleotide polymorphism of the 174G in the IL6 gene with ASPCR Mutation of interleukin IL6 SNP-screen-RV undiscovered-100 (Syntol, Russia); SYBR Green is an asymmetric cyanine dye used in molecular biology for staining nucleic acids (SY for Synthetic, BR for Bromide, Green for the fluorescent properties of the dye, which emits green light when excited) (Thermo Fisher Scientific, USA).

Equipment: the amplification reaction will be carried out using the following devices: Dtlight detecting amplifier according to TU 9443-003-96301278-2010, 4S1 modification and DTprime according to TU 9443-004-96301278-2010, 5M3 modification from NPO DNA Technology LLC (Protvino). The devices used ensure implementation of qualitative and quantitative studies using the allele-specific polymerase chain reaction (ASPCR) method without the stage of electrophoresis of PCR products in agarose gel using reagent kits based on the principles of fluorescence detection.

Analysis methods: detection of PCR results using intercalating agents (SYBR Green I with an emission wavelength of 520 nm). The amount of accumulated PCR amplification product is estimated directly during the temperature reaction cycles (real-time PCR). The quantitative analysis is based on the standard PCR curve study with an analysis of the accumulated fluorescent signal through the FAM channel using an appropriate mathematical apparatus.

As mentioned above, detailed presentation of information to each subject will allow to get rid of unnecessary questions, on the one hand, and to identify misunderstandings and address the issue locally to avoid the loss of a specific potential participant in the study, on the other hand.

To ensure a special solution of ethical and legal issues, each of the implications should be considered in a more detailed way.

INFORMED CONSENT

First and foremost, every patient is entitled to receive complete information about the purposes, methods and possible consequences of genetic testing. It is important to explain to the patient the significance of the results obtained, the risks and benefits of the procedure, as well as the degree of diagnostic accuracy. A person can make an informed decision regarding their participation in the study only after complete information has been obtained.

Moreover, informed consent should include an explanation of how the data obtained will be used. The patient should clearly understand who will get access to the results of the DNA analysis, who else can see the information, and what goals the researchers have.

CONFIDENTIALITY OF PERSONAL DATA

Another important implication is protection of personal information from unauthorized access by third parties. Genetic information is a valuable resource for pharmaceutical companies, insurance companies, and even employers. If such information gets into the hands of unauthorized persons, discrimination of patients can be possible, medical care or employment can be denied.

To prevent the leakage of confidential information, the following precautions should be observed:

- separate storage of biosamples and databases;
- limited access of laboratory staff to patient data;
- transferring encrypted data between medical institutions;
- regular security audit of information systems of medical institutions.

RISKS OF FALSE POSITIVE AND FALSE NEGATIVE RESULTS

No test warrants that the result can be totally accurate. False positive and false negative conclusions can significantly damage the patient's health, cause unnecessary concerns, or lead to unreasonable treatment costs.

Thus, it is important to repeat tests if there are doubts about accuracy of the initial result. It is necessary to clearly understand what limitations the method have and inform patients about all known risk factors associated with the test.

LEGAL RULES AND QUALITY STANDARDS

Special attention is given to development of laboratory research quality standards, regulation of laboratory activities and protection of rights of patients. In most countries, the legislation provides for mandatory licensing of genetic laboratories, certification of equipment and specialists working with biosamples.

For example, Federal Law No. 323-FZ "On the Basics of Public Health protection in the Russian Federation", adopted in 2011, establishes rules for handling personal data of citizens, prohibits collection and processing of biometric data without the consent of the subject, regulates the storage, processing and transfer of such information [8].

There are also international normative acts, such as the Convention for the Protection of Human Rights and Dignity of the Human Being with regard to the Application of Biology and Medicine (ETS No. 164), aimed at protecting the dignity of the individual and the right of every citizen to privacy and health [9].

It is the legislation that creates conditions for safe and effective introduction of new technologies into medical practice, ensuring compliance with the basic principles of medicine such as respect for patient autonomy, fair allocation of health resources and concern for well-being for all as a whole [10].

MODERN APPROACHES TO SOLVING ETHICAL PROBLEMS

The use of genetic research in clinical practice is accompanied by a number of serious ethical issues. Thus, to overcome the difficulties, a solution of the issue and a solution of the ethical implication are proposed (see Table).

Thus, according to the table, the most problematic ethical implications in genetic research include confidentiality and discrimination, which, most probably, consist of the fear of information leakage into the Internet and building further negative stereotypes about a person. It is the demonstration of the database formation that will help overcome these implications, where the subject can see how the identifying information is specifically deciphered. Thus, the procedure of virtual depersonalization is shown. Every patient should be adequately informed about the goals, methods, and potential risks of genetic testing. The doctor must provide all the details about the clinical significance of the results obtained, potential dangers and benefits of the procedure. Meanwhile, existing modern technologies can minimize many of the risks that arise during genetic research. For example, modern sequencers decode the sequence of nucleic acids with a high accuracy, which helps avoid errors while interpreting the analysis results.

Automated information management systems make database secure, prevent the leakage of confidential information and reduce the likelihood of abuse by clinic staff.

New methods used by the physician council ensure a rapid exchange of experience between different regions of the country and the world, improving the quality of medical care provided to the population.

If necessary precautions are taken, genetic research brings significant benefits to patients, doctors, and society as a whole. Such tests help identify pathologies in a timely manner, select the optimal treatment strategy and prevent complications.

CONCLUSION

Genetic research opens up new perspectives for prevention, diagnosis and therapy of various diseases of the musculoskeletal system. It also raises a number of complex ethical issues related to protection of privacy, compliance with the principle of voluntariness and equal distribution of health resources.

Compliance with strict rules for genetic research will make this tool an effective means of improving public health, minimizing negative consequences and protecting the interests of all participants in the process. Successful implementation of the project depends on competence of health professionals, public awareness and willingness of the state to support the initiatives of scientists and doctors.

Table. Stages of solving ethical issues

Stage	Issue and solution	Solution	Overcoming success* %
Awareness	The duty of total patient awareness of purposes, methods and risks of a genetic testing	To provide as many details about the essence and purpose of the study as possible	85
Confidentiality	Reliable storage and processing of data exclusively by authorized persons	To tell and, if possible, to show in what form and where the received data will be stored	60
Legal liability	Liability for non-compliance with regulatory requirements	To concentrate on elements of informed consent, where liability in case of non-compliance with the regulatory requirements of the researcher is provided	90
The risk of erroneous conclusions	Possible incorrect conclusions due to incorrect tests	Explain the importance of a repeated examination in case of a negative result that will help eliminate errors	95
Discrimination	Possible influence of genetic testing results on availability of services and insurance	Medical secrecy is above all, data is depersonalized and not included into registers	65

* Note: the issue solving success rate was obtained based on the present study on genetic polymorphism in trauma patients, where out of the initial 269 people, 61 refused to participate after the selection procedure according to the ethical implication criteria.

References

1. Visscher PM, Wray NR, Zhang Q, et al. Five Years of GWAS Discovery. Am J Hum Genet. 2021; 108(1): 7–18. DOI: 10.1016/j.ajhg.2020.11.007.
2. Savgachev VV, Yuriy AV, Shubin LB, et al. Baza dannykh algoritmov prognosticheskoy effektivnosti diagnostiki, lecheniya i profilaktiki na osnove geneticheskogo polimorfizma riskovoy patologii razvitiya oslozhneniy pri otkrytoy travme kostey goleni. Svidetel'stvo o registratsii bazy dannykh RU 2022621249, 30.05.2022. Zayavka № 2022621132 ot 20.05.2022. Russian.
3. Petrov SV, Smirnov AI. Gnoynyye oslozhneniya posle ortopedicheskikh operatsiy: patogenez i profilaktika. Travmatologii i Ortopedii im N. I. Priorova. 2020; 27(3):150–156. DOI: 10.18019/2311-2905-2020-27-3-150-156. Russian.
4. Wang Y, et al. Genetic Predisposition and Risk of Complications After Orthopedic Surgery: A Review. J Orthop Res. 2023: 41(4): 689–703. DOI: 10.1002/jor.25395.
5. Rushdy M, Elsayed MS, Ahmed R, Gaber AG, et al. IL-17A (rs2275913; G197A) gene polymorphism as predictor for disease severity and its correlation with IL-17 serum levels in COVID-19 patients. Egypt J Immunol. 2022; 29(3): 90–98.
6. Kirova TA. O pravovom znachenii meditsinskoy etiki. Ex jure. 2018; 1. Available from URL: https://cyberleninka.ru/article/n/o-pravovom-znachenii-meditsinskoy-etiki (accessed: 30.04.2025). Russian.
7. Beshir L. Research Ethics Committees in Laboratory Medicine. EJIFCC. 2020 Nov 20; 31(4): 282–291.
8. Donika AD, Kozhevnikov LL. Sootvetstviye norm Federal'nogo zakona «Ob osnovakh okhrany zdorov'ya grazhdan v Rossiyskoy Federatsii» normam i printsipam bioetiki. Bioetika. 2011; 2(8): 26–28. Russian.
9. Siluyanova IV. Problemy bioetiki v «konventsii o pravakh cheloveka i biomeditsine» (se,1997) i «osnovakh sotsial'noy kontseptsii» (rpts, 2000): sravnitel'nyy analiz. Bioetika. 2015; 8: (2): 28–32. Russian.
10. Yepifanova YeV, Chuprova AA, Khil' IM. K voprosu o sistematizatsii meditsinskogo zakonodatel'stva (teoretiko-pravovoy aspekt). Pravo i gosudarstvo: teoriya i praktika. 2019;12: 180. Russian.

Литература

1. Visscher PM, Wray NR, Zhang Q, et al. Five Years of GWAS Discovery. Am J Hum Genet. 2021; 108(1): 7–18. DOI: 10.1016/j.ajhg.2020.11.007.
2. Савгачев В. В., Юрий А. В., Шубин Л. Б. и др. База данных алгоритмов прогностической эффективности диагностики, лечения и профилактики на основе генетического полиморфизма рисковой патологии развития осложнений при открытой травме костей голени. Свидетельство о регистрации базы данных RU 2022621249, 30.05.2022. Заявка № 2022621132 от 20.05.2022.
3. Петров С. В., Смирнов А. И. Гнойные осложнения после ортопедических операций: патогенез и профилактика. Травматологии и ортопедии им Н. И. Приорова. 2020; 27(3):150–156. DOI: 10.18019/2311-2905-2020-27-3-150-156.
4. Wang Y, et al. Genetic Predisposition and Risk of Complications After Orthopedic Surgery: A Review. J Orthop Res. 2023: 41(4): 689–703. DOI: 10.1002/jor.25395.
5. Rushdy M, Elsayed MS, Ahmed R, Gaber AG, et al. IL-17A (rs2275913; G197A) gene polymorphism as predictor for disease severity and its correlation with IL-17 serum levels in COVID-19 patients. Egypt J Immunol. 2022; 29(3): 90–98.
6. Кирова Т. А. О правовом значении медицинской этики. Ex jure. 2018; 1. Режим доступа: [Электронный ресурс] URL: https://cyberleninka.ru/article/n/o-pravovom-znachenii-meditsinskoy-etiki (дата обращения: 30.04.2025).
7. Beshir L. Research Ethics Committees in Laboratory Medicine. EJIFCC. 2020 Nov 20; 31(4): 282–291.
8. Доника А. Д., Кожевников Л. Л. Соответствие норм Федерального закона «Об основах охраны здоровья граждан в Российской Федерации» нормам и принципам биоэтики. Биоэтика. 2011; 2(8): 26–28.
9. Силуянова И. В. Проблемы биоэтики в «конвенции о правах человека и биомедицине» (се,1997) и «основах социальной концепции» (рпц, 2000): сравнительный анализ. Биоэтика. 2015; 8: (2): 28–32.
10. Епифанова Е. В., Чупрова А. А., Хиль И. М. К вопросу о систематизации медицинского законодательства (теоретико-правовой аспект). Право и государство: теория и практика. 2019;12: 180.

RELIGION OR SCIENTIFIC RATIONALITY: SEARCH FOR ONTOLOGICAL FOUNDATIONS OF MEDICAL ETHICS

Kozlova OV ✉

Yaroslavl State Medical University, Yaroslavl, Russia

The article examines the relationship between religious values and rational principles of science. The purpose of the study is to identify common grounds of religion and scientific rationality to determine the foundations of medical ethics. The article examines the concept by Kavelin KD in relation to opinions of modern philosophers. Religion educates a moral person and provides guidelines for medical ethics, medicine and scientific knowledge, while science, rational knowledge, clarifies the general conditions of actual existence and provides a tool for arranging a human life. It is concluded that both religion and science display interest in the same task but in a different way. Thus, religion looks at the mental, subjective, and moral side, whereas science is interested in something external and objective.

Key words: medical ethics, religion, scientific rationality, society, personality, consciousness, nature, values

✉ **Correspondence should be addressed:** Olga V Kozlova
Kedrova St., 8, sq. 10, 150000, Yaroslavl, Russia; olgakozlova.1997@mail.ru

Received: 16.05.2025 **Accepted:** 03.06.2025 **Published online:** 30.06.2025

DOI: 10.24075/medet.2025.009

РЕЛИГИЯ ИЛИ НАУЧНАЯ РАЦИОНАЛЬНОСТЬ: ПОИСК ОНТОЛОГИЧЕСКИХ ОСНОВАНИЙ МЕДИЦИНСКОЙ ЭТИКИ

О. В. Козлова ✉

Ярославский государственный медицинский университет, Ярославль, Россия

В статье рассматривается соотношение религиозных ценностей с рациональными принципами науки. Цель исследования: выявить общие основания религии и научной рациональности для определения оснований медицинской этики. Исследуется концепция К. Д. Кавелина в соотношении с мнениями современных философов. Религия воспитывает нравственную личность и дает ориентиры медицинской этике, медицине и научному знанию, а наука, рациональное знание, выясняет общие условия действительного бытия и дает орудие для обустройства человеческой жизни. Делается вывод о том, что и религия, и наука подходят к одной и той же задаче с двух различных сторон: религия — с психической, субъективной, нравственной; наука — с внешней, объективной.

Ключевые слова: медицинская этика, религия, научная рациональность, общество, личность, сознание, природа, ценности

✉ **Для корреспонденции:** Ольга Валерьевна Козлова
ул. Кедрова, д. 8, кв. 10, 150000, г. Ярославль, Россия; olgakozlova.1997@mail.ru

Статья поступила: 16.05.2025 **Статья принята к печати:** 03.06.2025 **Опубликована онлайн:** 30.06.2025

DOI: 10.24075/medet.2025.009

Today, when rational thinking prevails, it is especially important to comprehend the true values of medical ethics. Thinkers of the past and present were interested how religious values and achievements of scientific rationality were related. Every person faces the problem of finding the true foundations for his being and for making the right decisions. “We all want to enjoy our lives, try to control it… Everyone is trying to make their small world most reliable and interesting. As a doctor, I have to confront a parallel reality such as pain, suffering and death on a daily basis”, writes Ariel Noltze, a German plastic surgeon [1].

It seems relevant to turn to works of Kavelin KD, an outstanding thinker of the 19th century, and review his position in discussion with modern philosophers. Traditionally opposed religious values and principles of scientific rationality are newly interpreted both in the philosophy of Kavelin KD and in the concepts of modern philosophers.

According to Kavelin’s concept, moral character and moral development of a human being can’t exist without free will or “without the opportunity, at own discretion and on own volition, to choose one way or another, to incline to one action or another, and to set a direction for an activity” [2]. The thinker believes that external circumstances can promote or prevent from implementation of certain human decisions. Thus, it can be concluded that struggle is a permanent law of a moralist. According to Kavelin, a person constantly struggles with the natural surroundings and himself to achieve his goals and to create a decent habitat in line with his goals. This is how a moralist grows up.

The thinker is convinced that science can determine the main goals in a human life. It is the positive sciences dealing with phenomena and external facts that have the task, firstly, to establish these phenomena and facts in their actual reality, and secondly, to determine the conditions that made these facts necessary and inevitable. As soon as these two tasks are completed, the work of science is over, since the phenomenon and the fact have been explained, the thinker says. Consequently, real science deals only with relative truths but not eternal ones. Therefore, Kavelin makes a conclusion that free will cannot be adopted by real science since freedom of choice rejects the necessity of phenomena. “If, at the discretion of the one who acts, a fact can happen one way or another, or even never happen at all, then there is no way to determine its law; real sciences, as we said, have the task to determine the laws of phenomena” [2].

Gerhard Medicus notes that with the lack of trust in empirical methods and "erroneous feelings," logical consistency and a scientist's own opinion are the only criteria for assessing credibility. "In this sense, philosophers and mathematicians are experts in evidence-based arguments and uncompromising demands for confidence" [3].

Kavelin KD states that as conclusions about mental life made by real sciences are used erroneously, personality and conditions of its activity have nothing to do with prevailing modern ideas: "real sciences perceive individuals only as parts of the final result to be considered and explained. The final result is a necessary phenomenon that follows certain laws; therefore, it does not depend on a personal activity, and therefore there is no need to take it into account" [2]. Gerhard Medicus notes that there is no use to assume that probabilistic processes are directly responsible for our free will at the level of reality described by quantum physics. "Conclusions about conscious processes based on quantum physics are questionable and can be compared with the equally doubtful act of opening a department of political science at the Institute of Biochemistry" [3].

According to Kavelin KD, personality is currently not perceived as a moral figure. Personalities are gradually turning into impersonal human units, deprived of a point of support in their moral existence and therefore easily replaceable. Meanwhile, dignity and moral character are formed under the influence of external life and activity as a member of the state and society. When evaluating a person, it is not internal motives, but the degree of necessity for society that matters, and it is external habits that will be taken into account to judge about a person. Fanaticism is another obsession, which is much more dangerous, says Caroline Emke. She notes the growing threatening global dynamics when people who believe differently or do not believe at all and who look differently from what is required by the approved standards are fundamentally rejected by the society. "This growing disapproval of any deviation is spreading and becoming more harmful. Because we, who are targeted by this hatred, usually lapse into silence in disgust allowing others to intimidate us because we can't resist this savagery and terror…" [4].

Kavelin is convinced that as long as the moral elements of personality remain neglected, these views will penetrate deeper into the minds of many educated people. The philosopher points out that it is in his modern era that everything possible is being done to meet human needs, while the human is becoming less and less able to use these benefits.

On the other hand, the thinker believes that the model when a person is separated from the rest of the world and finds his support in mental activity should have been replaced by another model with the central position in the research being occupied not by a single person but by the society. The world of knowledge and science opened up to a human owing to generalizations that could be implemented only through his communication with other people. As a person can develop and improve in society only, he should be viewed not as an independent unit, but as an integral part of the whole. In this regard, Kavelin warns against the misconception when people perceive a person as an integral part of an organism. "Since differentiation in humans is highly developed, a person in society occupies a more independent position and can go through a more intense individual development than the components of another living organism" [5]. The philosopher stresses that the ideal world helps a person leave the narrow circle of his personal existence and contemplate the universal.

Continuing Kavelin's thought, Costantino Esposito notes that our consciousness is mysterious. Consciousness always laughs at those who try to analyze it. "Since it is already a part of those who want to deconstruct it, deconstruction of consciousness is actually a proof of its existence" [6]. The philosopher wonders which of the realities is responsible for creating consciousness. However, it must be borne in mind that the person asking the question is already inside consciousness at this moment. Consciousness does not depend on our inability to explain its subjective manifestation. Costantino Esposito points out that the mystery of consciousness relates to human sensory and mental activity. We can say that consciousness is embodied in the feelings and thoughts of a person.

This raises an eternal question. How can a person have free will if he is determined by both external reality and consciousness? According to Kavelin, free will is not an illusion, but a real phenomenon. If everything in the world exists and happens under certain conditions and a human is an organic part of nature, it is difficult to imagine that a man has a force inside that creates phenomena beyond all conditions. According to the philosopher, everyone is meanwhile directly convinced that, under certain circumstances, people spontaneously and freely control their inner mood and external actions. Numerous observations have established differences between free actions and actions produced under pressure of passions, worries, and fear. The differentiation would not be possible if mental states were not based on freedom of mental activity.

In this regard, Jean-Baptiste Brenet notes that a person should then be a substance or a completely charged reality that does not need any support and exists independently of the existence or activity of other things in the world. But when a person is born, he is not the substance. "Mind is his entity; initially, the mind is just a preparation and capacity to abstraction of forms within the matter; this makes him subordinate to his body, feelings and imagination. A human being is a substance that only expects to be implemented as a promise that has to be kept and as luck that has to be experienced …" [7].

Chaadaev PY is convinced that the connection between moral phenomena is similar to one that unites physical phenomena; it is about continuity and succession. It is the effect of the moral phenomenon that promotes a person's self-development. "In the field of morality, people move forward not only for the pleasure of moving, there must also be a goal; a denied possibility of achieving perfection, which means reaching the goal, would simply make movement impossible" [8]. The thinker is convinced that people come into the world with a vague instinct for moral good. But this instinct can only be fully implemented in a more complete idea of ethics, which develops throughout life.

Therefore, Patricia Churchill notes that cultivating virtues such as compassion and honesty is beneficial. If these virtues turn into habits, they will guide the process of fulfilling limiting conditions towards making decisions that are morally acceptable to a person. Thus, being fixed in consciousness, these habits allow the brain not to calculate and evaluate from scratch all the facts that influence the choice: "… if you are used to showing, say, kindness and responsiveness to everyone around you, you will not have to waste your time and effort thinking about what to do in a standard everyday situation. In case of an extraordinary event, a conditional habit can be useful" [9]. The researcher notes that the utilitarian's systemic brain that produces a line of decisions has to spend so much extra effort deciding

whether an action was correct that one can only wonder if such a person is capable of making decisions at all and completing at least one task.

Kavelin KD stresses that the undoubted connection between a person and the world around him that forces development of knowledge and science does not solve all the problems that arise in the course of historical development. According to the thinker's concept, a person with his inner world and its secrets is not a continuation, clarification and addition of the surrounding world, but, on the contrary, the person denies it, "escapes from its evils and sufferings, seeks support in himself and in the name of it seeks to recreate the entire real world and conditions of its existence" [5].

Thus, Kavelin asks what this hostile opposition of the inner world to the outer world means and why it is the final act of the epochs of cultures and civilizations and not their very beginning. According to the philosopher, late protests in the name of inner and spiritual peace against the environment is easily explained by the law of differentiation, which is equally noticeable when nature and man develop. The thinker notes that the unity not seen in germ is so much separated during subsequent growth that it is difficult to find and determine the mutual connection between previous parts of the whole.

In this regard, Michael Marder writes that the consequences of metaphysics criticism are not entirely negative, given that the contours of life, such as plants, become visible as a result of hermeneutical multiplication of its meanings, freed from the reductive tendencies of metaphysics. Positive dimension of plant existence as a consequence of metaphysics criticism, leads to inversion of traditional values, putting one above the other. "More importantly, it covers key existential attributes that philosophers, as a rule, saved for humans only" [10]. Michael Marder calls his concept 'vegetative existentialism' and states that it would be wrong to insist on traditional metaphysical separation of a spirit from a body though this is one of many dichotomies of oneself and the other person, depth and surface, life and death, the whole and a great many.

Kavelin KD draws attention to the fact that all real life is a struggle. Everything that exists, from lower to higher organisms, lives at the expense of one another, conquering its existence and constantly being in danger of becoming a victim. The philosopher points out that the general law of life is most clearly applied to a man who is the most developed and complex of all organisms. Man is constantly fighting nature, people similar to him, and society, now defeating them, then being defeated by someone more superior.

In this regard, Jean-Pierre Dupuy states that the current discussion about the changing attitude towards nature caused by new technologies boils down to the fact that "deep ecology" presents nature as an unshakable example of balance and harmony. As a result, man appears to be an irresponsible and dangerous predator. Therefore, the goal of the whole project of modern humanism is to take man out of nature and turn him into the ruler of the world and himself, points out Jean-Pierre Dupuy. "The metaphysics in question definitely insists on its monism: it is no longer claimed today that everything in the world originates from one substance, but that everything — nature, life, consciousness — is subject to general principles of organization" [11]. Therefore, he comes to the conclusion that the motto "naturalize consciousness" becomes the goal of cognitive sciences. These sciences should newly supply consciousness with its rightful place in nature, concludes Jean-Pierre Dupuis.

Based on this, we ask how to eliminate the contradiction between the unity of all that exists and continuous struggle of this existing world with itself. To answer this question, Kavelin compares science and religion. According to the philosopher, the task of science is to know the laws and necessary conditions of phenomena. For religion, it is not the objective truth, but a person's parting words to a spiritual and moral life that matters the most. Religion rejects anything that does not comply with the goal as harmful and evil. "Vigilantly and jealously protecting only personal, individual spiritual and moral existence, it stops attempts of knowledge to penetrate the mysteries of existence where they could shake the foundations of personal spiritual life, stating that these conditions are hidden from human knowledge and incomprehensible to the mind" [5].

According to Kavelin's concept, the religious worldview is based on the sole basic idea — to preserve, guide and educate a person spiritually and morally, to support his soul against temptations on the path of life. Therefore, the philosopher believes that everything is adjusted to this life goal: different branches of art, philosophy, forms of life, and society. The power of religion and a mystery of its enormous influence on people consist in concern for satisfaction of spiritual needs of the individual's existence.

According to Kavelin KD, the tasks of knowledge are completely different and its methods differ from those necessary for spiritual and moral education of an individual. Science studies not the subject as a whole, in living reality, but conditions and laws of its existence and activity, and also transforms these laws into general formulas of being. The thinker notes that science always begins with decomposition of living reality into components and as a result gets something completely different from the reality that underwent scientific analysis. This result consists in discovering and understanding the conditions and laws of real life. That is why conclusions of science represent completely different combinations than those that actually exist. Science gives us knowledge of what exists, but from a special point of view, from a known side which is abstracted from reality and combines in our mind in a completely different way, Kavelin concludes.

According to Costica Bradatan, accepting the definition of science as a self-transforming practice makes people absolutely vulnerable. Costica Bradatan compares the philosopher with a tightrope walker who performs without insurance because he eternally balances between adjustment to the demands of the world and his own ethical principles. "It happens because for such thinkers, philosophy is not just a set of doctrines that can be ignored or discarded if needed. It is a way of life that goes through your entire biography, and this choice is of a significant existential nature" [12].

Another question is why we need the new combinations and transformed reality. In real life, the role of these ones and thus of a science is essential. According to Kavelin, due to new combinations of reality and owing to their help, a human can produce what is useful and necessary and discard what is useless and unnecessary. The philosopher states that until now nobody could ever define the boundaries of knowledge and where it should stop. "Like a large mass of snow rolling down a mountain, knowledge, as it develops, does not only get increased but also expands its tools and turns into a huge power. But we cannot and must not request from

knowledge something that it cannot provide us with due to its nature" [5].

Kavelin KD defines the mind as a special process, a special function of a human being and of human nature. Mind creates nothing, it can only produce new combinations of something that already exists, and it is always generalized. The generalized statements possess no reality outside a human, though philosophy has claimed the opposite for a while. The thinker believes that the only reason for the mistake was that mind as an organic property acts subconsciously, against our will and beyond our comprehension. According to Ian Tattersall and Rob DeSalle, human beings will always be an unsolved mystery. It is ironic that people belong to the only species in the world who can look back and examine themselves. So, they are mysterious to themselves only. "Our unique cognitive style separates us from the rest of nature not only because of our ability to understand and master the world we live in, but also because of our ability to make up reductionist stories about it and believe them" [13].

As Kavelin KD states, we must recognize that knowledge is nothing more than a special way of relating to the world around us and ourselves, which is unique to the human race and serves as a tool to achieve its goals. A human has a goal inside. It is about how to satisfy various human needs. The philosophers says that we refer to knowledge as a tool because the knowledge itself, in a theoretical and pure sense, is the same reality transformed by the mental process and presented to consciousness. Thus, living reality is still a dead and dry abstraction from true foundations of objective reality.

Jean Baudrillard, who continued Kavelin's thought, wrote that the universe is not in equilibrium but rather operates on extremes, it is rather about radical antagonism than about synthesis. Thus, antagonism is present. It has an ecstatic form of a pure object and winning strategy of the object in relation to the subject. "We will not look for a change and oppose something steady and something changeable, we will find something more changeable than changeability itself; it will be metamorphosis."… We will not distinguish the true from the false, we will find something more deceptive than deception itself; it will be illusion and appearance…" [14].

So, Kavelin insists that the only reason for their significance difference can be associated with various purposes, role and tasks of religion and science. Religion educates a moral person and provides guidelines for medical ethics, medicine and scientific knowledge, while science, rational knowledge, clarifies the general conditions of actual existence and provides a tool for arranging a human life.

Thus, both religion and science display interest in the same task but in a different way. Religion looks at the mental, subjective, and moral side, whereas science is interested in something external and objective. According to Kavelin, they are opposed only because of deep misunderstanding and unclear perception of their mutual relations, circle and boundaries of their activity. "The purpose of religion is not knowledge; therefore, it should not be opposed to it and act as its enemy, no matter what results and conclusions it may come to. Knowledge, just like science, should not act against religion, as it is aimed not at moral upbringing of humans, but at discovery of general conditions and laws of existence" [5].

Kavelin KD is convinced that the main goal of the society that unites people with religious and scientific background is to detect average terms of their peaceful and harmless coexistence. The goal can be achieved when the reasons for the opposites will be comprehended by both parties and when they both will voluntarily outline the sphere of their activity. At the same time, it should be borne in mind that knowledge and science do not provide an accurate answer to many questions; in practical application, we can only approach an impossible ideal and try to be satisfied with that.

In this regard, the ideas of Kavelin KD, an outstanding Russian philosopher, can be considered as a spiritual testament to the modern generation. Indeed, as Geert Lovink accurately noted, "we need to extend deconstruction of the Western subject to the non-human agency of the Internet … Only then can we understand cultural policy in a clearer way" [15]. Geert Lovink states that to avoid distraction, it is necessary to find a new way of thinking, which could be useful for 'post-digital era' and which could admit that the Internet would neither disappear nor become an obstacle; instead, it will promote spiritual growth of a person.

References

1. Nol'tse Ariel'. Operatsiya «Chelovek». Skal'pel' i krest — dva ostriya dlya novogo nachala. Per. s nem. I. Lobanova. Zaokskiy. Istochnik zhizni. 2020; 7–8. Russian.
2. Kavelin K. D. Sobr. soch. Spb. 1899; 3: 383–385. Russian.
3. Medikus Gerkhard. Byt' chelovekom: Preodoleniye razryva mezhdu naukami o tele i naukami o dushe. Per. A. B. Kutlumuratov. M. Izdatel'skiy Dom YASK. 2020; 50–51. Russian.
4. Emke Karolin. Protiv nenavisti. Per. s nem. Anny Kukes. Spb. Izdatel'stvo Ivana Limbakha. 2024; 17. 5. Kavelin K. D. Nash umstvennyy stroy. St. po filosofii, russkoy istorii i kul'ture. M. Pravda. 1989; 351–521. Russian.
5. Kavelin KD. Nash umstvennyy stroy. St. po filosofii, russkoy istorii i kul'ture. M. Pravda. 1989; 351–521.
6. Espozito Konstantino. Sovremennyy nigilizm. Khronika. Per. s ital. Bogdanovoy YA.A. M. RIPOL klassik 2022; 74. Russian.
7. Brene Zhan-Batist. Chto znachit myslit'? Arabo-latinskiy otvet. Per. s frants. Dar'ya Shakhova. M. Ad Marginem Press. 2024;109. Russian.
8. Chaadayev P.YA. Poln. SOBR. soch. i izbrannyye pis'ma. V 2-kh t. M. Nauka. 1991; 1:472. Russian.
9. Cherchlend P. Sovest'. Proiskhozhdeniye nravstvennoy intuitsii. Per. s angl. Mariya Desyatova. M. Al'pina non-fikshn. 2021; 213. Russian.
10. Marder Maykl. Rastitel'noye myshleniye. Filosofiya vegetativnoy zhizni. Per. s angl. Denis Shalaginov. M. Ad Marginem Press. 2024; 27. Russian.
11. Dyupyui ZH.-P. Znak svyashchennogo. Per. s frants. A. Zakharevich. Pod red. A. Grinbauma. M. Novoye literaturnoye obozreniye. 2021; 92. Russian.
12. Bradatan K. Umiraya za idei. Ob opasnoy zhizni filosofov. Per. s angl. Ye. V. Muzykina. M. Novoye literaturnoye obozreniye. 2024; 25. Russian.
13. Tatterson I., De Sall' R. Chelovek neozhidannyy: genetika, povedeniye i svoboda vybora. Per. s angl. N. Pakhmutova. M. Izdatel'skiy Dom YASK. 2023; 11. Russian.
14. Bodriyyar ZH. Fatal'nyye strategii. Per. s fr. A. Kachalova. M. Gruppa kompaniy «RIPOL klassik». «Pangloss». 2019; 8. Russian.
15. Lovink Gert. Kritecheskaya teoriya Interneta. Per. s angl. Dmitriy Lebedev, Petr Torkanovskiy. M. Ad Marginem Press, 2024; 31. Russian.

Литература

1. Нольце Ариэль. Операция «Человек». Скальпель и крест — два острия для нового начала. Пер. с нем. И. Лобанова. Заокский. Источник жизни. 2020; 7–8
2. Кавелин К. Д. Собр. соч. Спб. 1899; 3: 383–385.
3. Медикус Герхард. Быть человеком: Преодоление разрыва между науками о теле и науками о душе. Пер. А. Б. Кутлумуратов. М. Издательский Дом ЯСК. 2020; 50–51.
4. Эмке Каролин. Против ненависти. Пер. с нем. Анны Кукес. Спб. Издательство Ивана Лимбаха. 2024; 17.
5. Кавелин К. Д. Наш умственный строй. Ст. по философии, русской истории и культуре. М. Правда. 1989; 351–521.
6. Эспозито Константино. Современный нигилизм. Хроника. Пер. с итал. Богдановой Я. А. М. РИПОЛ классик 2022; 74.
7. Брене Жан-Батист. Что значит мыслить? Арабо-латинский ответ. Пер. с франц. Дарья Шахова. М. Ад Маргинем Пресс. 2024;109.
8. Чаадаев П. Я. Полн. СОБР. соч. и избранные письма. В 2-х т. М. Наука. 1991; 1:472.
9. Черчленд П. Совесть. Происхождение нравственной интуиции. Пер. с англ. Мария Десятова. М. Альпина нон-фикшн. 2021; 213.
10. Мардер Майкл. Растительное мышление. Философия вегетативной жизни. Пер. с англ. Денис Шалагинов. М. Ад Маргинем Пресс. 2024; 27.
11. Дюпюи Ж.-П. Знак священного. Пер. с франц. А. Захаревич. Под ред. А. Гринбаума. М. Новое литературное обозрение. 2021; 92.
12. Брадатан К. Умирая за идеи. Об опасной жизни философов. Пер. с англ. Е. В. Музыкина. М. Новое литературное обозрение. 2024; 25.
13. Таттерсон И., Де Салль Р. Человек неожиданный: генетика, поведение и свобода выбора. Пер. с англ. Н. Пахмутова. М. Издательский Дом ЯСК. 2023; 11.
14. Бодрийяр Ж. Фатальные стратегии. Пер. с фр. А. Качалова. М. Группа компаний «РИПОЛ классик». «Панглосс». 2019; 8.
15. Ловинк Герт. Критеческая теория Интернета. Пер. с англ. Дмитрий Лебедев, Петр Торкановский. М. Ад Маргинем Пресс. 2024; 31.

ETHICS IN MEDICAL RESEARCH AND PUBLICATIONS

Pleshchev IE[1] ✉, Shishkin AA[2], Ivashkovskaya AV[3]

[1] Yaroslavl State Medical University, Yaroslavl, Russia

[2] Pirogov Russian National Research Medical University, Moscow, Russia

[3] Yaroslavl State University named after Demidov PG, Yaroslavl, Russia

The purpose of this study is to examine the existing ethical standards and guidelines to provide a comprehensive overview of ethical issues and processes related to research and publications in domestic and international medical practice. The importance of informed consent, data integrity, plagiarism, authorship disputes, and conflicts of interest are just some of the key topics briefly covered in the article. It is obvious that ethical standards and regulations in medical (clinical) research are crucial for determining how research is conducted and how scientific articles are published.

Keywords: medical ethics, research, publications, medical law, academic dishonesty

Author contribution: Pleshchev IE — the concept and design of the study, responsibility for the article integrity, writing the text; Shishkin AA, Ivashkovskaya AV — writing and editing the text.

✉ **Correspondence should be addressed:** Igor E. Pleshchev
Revolutsionnaya St., 5, Yaroslavl, 150000, Russia; doctor.pleshyov@gmail.com

Received: 30.04.2025 **Accepted:** 17.05.2025 **Published online:** 30.05.2025

DOI: 10.24075/medet.2025.004

ЭТИКА В МЕДИЦИНСКИХ ИССЛЕДОВАНИЯХ И ПУБЛИКАЦИЯХ

И. Е. Плещёв[1] ✉, А. А. Шишкин[2], А. В. Ивашковская[3]

[1] Ярославский государственный медицинский университет, Ярославль, Россия

[2] Российский национальный исследовательский медицинский университет имени Н. И. Пирогова, Москва, Россия

[3] Ярославский государственный университет им. П. Г. Демидова, Ярославль, Россия

Целью данного исследования является изучение существующих этических норм и руководящих принципов для предоставления всестороннего обзора этических вопросов и процессов, связанных с исследованиями и публикациями в отечественной и международной медицинской практике. Важность информированного согласия, целостности данных, наличие плагиата, споров об авторстве и конфликтов интересов — вот лишь некоторые из ключевых тем, кратко затронутых в статье. Очевидно, что этические нормы и правила в медицинских (клинических) исследованиях, имеют решающее значение для определения того, как проводятся исследования и как публикуются научные статьи.

Ключевые слова: врачебная этика, исследования, публикации, медицинское право, академическая недобросовестность

Вклад авторов: Плещёв И. Е. — концепция и дизайн исследования, ответственность за целостность всех частей статьи, написание текста; Шишкин А. А., Ивашковская А. В. — написание текста, редактирование.

✉ **Для корреспонденции:** Плещёв Игорь Евгеньевич
ул. Революционная, д. 5, г. Ярославль, 150000, Россия; doctor.pleshyov@gmail.com

Статья поступила: 30.04.2025 **Статья принята к печати:** 17.05.2025 **Опубликована онлайн:** 30.05.2025

DOI: 10.24075/medet.2025.004

Dictionary of the Russian language by Ozhegov SI and Shvedov NYu (1997) states that ethics is the philosophical study of moral phenomena, its development, principles, norms, and role in the society [1, 2]. In turn, the Oxford Dictionary (1989) provides a similar definition of ethics as moral principles that govern human behavior, or a whole system of moral principles and rules of conduct [3]. As the definitions state, ethics includes people and their actions. In biomedical fields, the people involved in research are researchers, whereas the subjects they research and the actions involved are associated with how the researchers design, execute, analyze, report, and distribute research results to colleagues and general public for benefit [4].

Researchers execute various roles as team leaders, team members, authors, contributors, reviewers, and editors, and interact with publishers at different points of their research activity. Other people, fabrics, materials, machines and devices, and/or software (including artificial intelligence, robots, and ChatGPT) can be subjects of research [5].

Thus, ethics in research means a complex interaction of researchers, subjects, devices, authors, editors, reviewers, and publishers. All of them have their own role in obtaining the research result. Ethics in research is an essential element of the scientific process and is crucial for integrity and reliability of the research results.

FORMATION OF ETHICS IN MEDICAL RESEARCH

Ethics in research originates from the Second World War. The Nuremberg Code was developed in August 1947 as a result of the horrors that Nazi doctors committed in the name of scientific research. The Code provided a list of moral principles or ethical principles for conducting medical experiments on humans. These included the need for a preliminary animal experiment as the basis for research, voluntary informed consent of people to participate in the research, and avoidance of unnecessary suffering of test subjects.

At the 18th General Assembly in 1964, the World Medical Association (WMA) developed the Helsinki Declaration, which sets out guidelines for medical research involving human participants as subjects. It was amended nine times, most recently at the

64th General Assembly in Brazil in October 2013 [6]. In the United States, the Belmont Report (1979) identified three basic principles that guide research involving people as subjects, which include respect for the individual, charity, and justice. Respect for the individual requires that researchers treat research participants as people who have a right to make decisions about their own lives [7]. It includes voluntary participation, informed consent, confidentiality, and well-being of participants.

Codes of research ethics are a set of principles, guidelines, and standards that provide guidance for research ethical and responsible conduct. Various guiding principles for ethical research have been developed at the international and national levels, including the ones from the World Health Organization (WHO), the International Council for Harmonization (ICH, 1990), and the Ministry of Health of the Russian Federation. They were as follows: the Code of Professional Medical Ethics of the Russian Federation (2012), the Procedure for Organizing and Conducting Ethical Expertise (2024), guidelines of the Ministry of Health and the United States Department of Human Services (HHS) and the Indian Medical Research Council (ICMR), etc. The essence of these guidelines is to protect the interests of research subjects, especially vulnerable groups, minimize harm and risk, and regulate research in order to conduct legitimate and high-quality research [4].

ETHICAL RESEARCH IN RUSSIA

In Russia, the Ethics Committee of the Ministry of Health of the Russian Federation, a permanent body, has been established to protect the life, health and rights of patients who receive medical care as part of clinical testing of prevention, diagnosis, treatment and rehabilitation methods, as well as to review clinical testing protocols [8]. The Committee is guided by the Constitution of the Russian Federation, federal constitutional laws, and orders of the President and Government of the Russian Federation.

Local ethics committees (LECs) have been established in Russian universities and research institutes to monitor the quality of research and protect patients' rights. All research proposals in biomedical, behavioral sciences, or social research involving humans as subjects, including their biological materials and related data, must be reviewed and approved by LECs prior to starting the project. These committees should follow the guiding principles of the Helsinki Declaration, other relevant ethical codes and orders of local and international law.

In all higher medical institutions of Russia, "bioethics" is a mandatory discipline, where the search for options and solutions to controversial topics of Russian society in the field of medicine, law, ethics and science are discussed [9]. Some universities, such as MGIMO, offer educational programs in Publication Culture and Research Ethics. The Institute of Psychology, Sociology and Bioethics has been established at YSMU (Yaroslavl State Medical University), which is a modern scientific and educational platform for these programs. At Pirogov Russian National Research Medical University, Bioethics and Legislation in Biomedicine is a mandatory part of Biomedicine.

Training in Bioethics is currently carried out in 44 universities of Russia.

ETHICS OF SCIENTIFIC PUBLICATIONS

To prepare a high-quality publication, researchers must follow reporting criteria for each research type. Guidelines published for each type of research are available. The roles involved in the publishing process include authors, reviewers, editors, and publishers, and each role has important contributions and responsibilities in the publishing process. Moreover, there are organizations that develop guidelines for maintaining ethics in publications and journals. The most notable ones include the International Committee of Medical Journal Editors (ICMJE), the Committee on Publication Ethics (COPE) and the World Association of Medical Editors (WAME). These organizations regularly develop guidelines on various issues related to each aspect of the publishing process, including developing ethics for authors, reviewers, editors, and publishers [7].

Ethical issues related to authors include authorship, plagiarism, fabrication and falsification of data, conflicts of interest, and data transparency [10]. The author comes into play when the research is completed and the manuscript starts being prepared. The research material is published when a suitable journal for publication is found and the manuscript is prepared according to the journal recommendations. Then the manuscript is sent to the editorial office of the journal, and we are patiently waiting the review results. It is necessary to be polite while communicating with the journal. When the manuscript is sent back for revision, the corresponding author must answer all the questions of the reviewers and editors, politely addressing them during the discussion. If the authors disagree with the reviewers, they should explain their position in a polite and reasonable way. It should be remembered that reviewers are experts in the field who have found time in their busy schedules to review and improve the manuscript. The corrected manuscript is being sent, and the decision of the editorial board of the journal is being awaited. After accepting the article, the author responsible for correspondence must promptly respond to the issuing (technical) editor of the journal to approve the proofreading before it is sent for publication. Since copyrights are transferred to the journal, the rights to distribute the article must be executed in accordance with the journal terms. The author must be aware of his rights regarding distribution of the published content [11].

In turn, the reviewer has a number of serious obligations during the review of a scientific research article. In each publication, checklists should be published for reviewers to evaluate the manuscripts including:

1) Compliance with the approved deadlines specified by the journal;
2) Providing an unbiased assessment of the manuscript;
3) Suggesting useful criticism to improve the article;
4) Compliance with ethical standards to ensure that the research was done in accordance with ethical standards, and the results are presented truthfully and accurately;
5) Confidentiality (the manuscript and its contents should not be disclosed to anyone who is not involved in the review process).

The scientific editor of the journal should not only communicate with the authors and control the review process, but also resolve any issues arising from misconduct of the authors or reviewers, as well as making appropriate decisions regarding a specific issue based on ethical principles.

CONSEQUENCES OF ETHICAL VIOLATIONS

Research misconduct is a serious problem. It can lead to misleading results that may affect scientific research and further research on the topic in the future. Misconduct is fabrication and falsification of data (deceptive use of statistics), or plagiarism during the proposal, execution, review, or reporting of a study [12, 13].

Fabrication and falsification of data directly threaten the goals of science, as such behavior leads to publication of erroneous results, which undermines the search for knowledge and truth. Although plagiarism does not imply publication of erroneous results, it indirectly threatens the goals of science, as it is a form of intellectual theft that negatively affects the social structure of science, undermining trust among researchers, creating hostility and resentment, and hindering career growth [7, 13].

Misconduct is any highly unethical behavior that threatens the integrity of science and can be clearly identified.

Conclusions. Ethics in medical research and publications plays a crucial role in establishing the authority and standard of scientific work. This study focuses on the key concepts of ethics that guide the publication process and various types of research. It also highlights the need for frameworks and guidelines for specific medical fields of clinical research.

References

1. Ozhegov SI, Shvedova NYu. Tolkovyy slovar' russkogo yazyka. Rossiyskaya akademiya nauk. Institut russkogo yazyka imeni VV Vinogradova. 4-ye izd. Dop. M.: Azbukovnik, 1997; 3(5): 944 Russian.
2. Ozhegov SI. Tolkovyy slovar' russkogo yazyka: okolo 100000 slov, terminov i frazeologicheskikh vyrazheniy. Pod red. prof. LI Skvortsova. M.: Mir i obrazovaniye, 2018; 736 s. Russian.
3. Oxford English Dictionary, second edition, edited by John Simpson and Edmund Weiner. Clarendon Press. Twenty volumes. 1989.
4. Roy S, Shah MH, Ahluwalia A, & Harky A. Analyzing the Evolution of Medical Ethics Education: A Bibliometric Analysis of the Top 100 Cited Articles. Cureus. 2023;15(7): e41411. DOI: 10.7759/cureus.41411.
5. Barrington NM, Gupta N, Musmar B, Doyle D, Panico N, Godbole N, Reardon T, & D'Amico RS. A Bibliometric Analysis of the Rise of ChatGPT in Medical Research. Medical sciences (Basel, Switzerland). 2023; 11(3): 61. DOI: 10.3390/medsci11030061.
6. Andersson H, Svensson A, Frank C, Rantala A, Holmberg M, & Bremer A. Ethics education to support ethical competence learning in healthcare: an integrative systematic review. BMC medical ethics. 2022; 23(1): 29. DOI: 10.1186/s12910-022-00766-z.
7. Kambhampati SBS, Menon J, Maini L. Ethics in Research and Publications. Indian J Orthop. 2023; 57(11): 1722–1734. Published 2023 Aug 28. DOI:10.1007/s43465-023-00971-x.
8. Prikaz Minzdrava Rossii ot 10.07.2015 № 435n «Ob Eticheskom komitete Ministerstva zdravookhraneniya Rossiyskoy Federatsii». Byulleten' normativnykh aktov federal'nykh organov ispolnitel'noy vlasti. 2015; 42. Russian.
9. Firsova OA, Pantyukhin DV. Prepodavaniye distsipliny «bioetika» i nravstvennoye stanovleniye lichnosti studentov meditsinskogo instituta. Uchenyye zapiski OGU. Seriya: Gumanitarnyye i sotsial'nyye nauki. 2021; 93: 4. Russian.
10. Habibzadeh F. Plagiarism: A Bird's Eye View. J Korean Med Sci. 2023; 38(45): e373. Published 2023 Nov 20. DOI: 10.3346/jkms.2023.38.e373.
11. Akbulut S, Sahin TT. Recommendations for the ethical guidelines for publication of scientific studies: The responsibilities of editors, reviewers and the authors. Ann Med Surg (Lond). 2021; 72: 103047. Published 2021 Nov 12. DOI: 10.1016/j.amsu.2021.103047.
12. Rashbi ND. Publikatsionnaya etika: moral'nyye printsipy i kul'turnyy dissonans. Nauchnyy redaktor i izdatel'. 2017; 2(2–4): 107–112. DOI: 10.24069/2542-0267-2017-2-4-107-112. Russian.
13. Resnik DB. Is it time to revise the definition of research misconduct? Account Res. 2019; 26(2): 123–137. DOI: 10.1080/08989621.2019.1570156.

Литература

1. Ожегов С. И., Шведова Н. Ю. Толковый словарь русского языка. Российская академия наук. Институт русского языка имени В. В. Виноградова. 4-е изд., доп. М.: Азбуковник, 1997; 3(5): 944.
2. Ожегов С. И. Толковый словарь русского языка: около 100000 слов, терминов и фразеологических выражений. Под ред. проф. Л. И. Скворцова. М.: Мир и образование. 2018; 736 с.
3. Oxford English Dictionary, second edition, edited by John Simpson and Edmund Weiner. Clarendon Press. Twenty volumes. 1989.
4. Roy S, Shah MH, Ahluwalia A, & Harky A. Analyzing the Evolution of Medical Ethics Education: A Bibliometric Analysis of the Top 100 Cited Articles. Cureus. 2023; 15(7): e41411. DOI: 10.7759/cureus.41411.
5. Barrington NM, Gupta N, Musmar B, Doyle D, Panico N, Godbole N, Reardon T, & D'Amico RS. A Bibliometric Analysis of the Rise of ChatGPT in Medical Research. Medical sciences (Basel, Switzerland). 2023; 11(3): 61. DOI: 10.3390/medsci11030061.
6. Andersson H, Svensson A, Frank C, Rantala A, Holmberg M, & Bremer A. Ethics education to support ethical competence learning in healthcare: an integrative systematic review. BMC medical ethics. 2022; 23(1): 29. DOI: 10.1186/s12910-022-00766-z.
7. Kambhampati SBS, Menon J, Maini L. Ethics in Research and Publications. Indian J Orthop. 2023; 57(11): 1722–1734. Published 2023 AugE; 28. DOI:10.1007/s43465-023-00971-x.
8. Приказ Минздрава России от 10.07.2015 № 435н «Об Этическом комитете Министерства здравоохранения Российской Федерации». Бюллетень нормативных актов федеральных органов исполнительной власти. 2015; 42.
9. Фирсова О. А., Пантюхин Д. В. Преподавание дисциплины «биоэтика» и нравственное становление личности студентов медицинского института. Ученые записки ОГУ. Серия: Гуманитарные и социальные науки. 2021; 93: 4.
10. Habibzadeh F. Plagiarism: A Bird's Eye View. J Korean Med Sci. 2023;38(45): e373. Published 2023 Nov 20. DOI: 10.3346/jkms.2023.38. e373.
11. Akbulut S, Sahin TT. Recommendations for the ethical guidelines for publication of scientific studies: The responsibilities of editors, reviewers and the authors. Ann Med Surg (Lond). 2021; 72: 103047. Published 2021 Nov 12. DOI: 10.1016/j.amsu.2021.103047.
12. Рашби Н. Д. Публикационная этика: моральные принципы и культурный диссонанс. Научный редактор и издатель. 2017; 2(2–4): 107–112. DOI: 10.24069/2542-0267-2017-2-4-107-112.
13. Resnik DB. Is it time to revise the definition of research misconduct? Account Res. 2019; 26(2): 123–137. DOI: 10.1080/08989621.2019.1570156