THE WHITE PAPER

of ethics in **artificial intelligence**

AI ALLIANCE RUSSIA

Moscow
2024

42

answers
to popular
questions
about AI

The team of authors edited by A. V. Neznamov

The White Paper of ethics in the field of artificial intelligence. The 2024 edition.

Complied by A. V. Neznamov

Scientific editor A. G. Krainov

Editors: S. A. Fokina, I. A. Cheremnykh

ISBN 978-5-6052008-7-1

The White Paper on Ethics in AI is a detailed and carefully structured review of the most pressing ethical issues that arise in the development of artificial intelligence. The focus is on 42 key issues that cover a wide range of topics, from data privacy and autonomous solutions to the creation of digital imitations and responsibility for the use of AI. The authors offer readers an in-depth analysis of modern AI technologies and reveal how their development affects moral and social aspects.

This book does not seek to provide definitive answers — rather, it opens up a space for dialogue, where each of the questions becomes a starting point for reflection and discussion. The publication is aimed at researchers, teachers, psychologists, medical workers, lawyers, developers and anyone who cares about how AI is changing our world and what it means for the future of each of us.

ISBN 978-5-6052008-7-1

Table of contents

About the book 6

Methodology 8

Chapter 1. Ten most popular ethical issues related to the development of AI 16

01. The ‘trolley problem’: what choice should an unmanned vehicle make over human life in the event of an inevitable collision? 18
02. The problem of digital human imitations: is it acceptable to create them? 23
03. The ‘black box problem’: is it possible to understand the principles of AI system operation and explain them to the user? 28
04. The debate over the need to inform: should people always be made aware that they are interacting with AI? 33
05. The problem of job cuts: will mass introduction of AI lead to people losing their jobs? 38
06. The challenge problem: should a person always be able to challenge a decision made with use of AI? 43
07. The problem of AI bias: is it possible to solve it? 48
08. The problem of accountability: using the example of medicine, what responsibility does an AI developer have in case of harm to a patient's health? 52
09. The problem of delegating decision-making: in the case of the judiciary, will AI be able to replace a judge? . . . 57
10. The problem of social rating: is it ethical to use AI to create a social rating? 62

Chapter 02. AI and Confidentiality 66

11. Is it ethical to use personal data for AI learning? 68
12. Is it ethical to collect user data from a smartphone or smart device for AI training? 71
13. Is it ethical to use AI in mass video surveillance? 74
14. Is it ethical to use AI to predict and prevent crimes? 77
15. Is it ethical to use AI for scoring in retail, finance and other specific applications? 80

Chapter 03. AI and integrity 84

16. The learning challenge: how to avoid AI learning based on false information? 86
17. The problem of spreading malicious or misleading information using AI: how to mitigate this? 89

18. Is it ethical for algorithms to offer the user goods and services that do not correspond to their usual preferences? 92

Chapter 04. Generative AI 96
19. Is it possible to trust information obtained with the help of generative AI and AI-based search engines? 98
20. Is it ethical to synthesize human speech using AI? 101
21. Is it ethical to use generative AI in art and design? 104
22. Is it ethical not to indicate that content has been generated with use of AI? 107
23. Will generative AI affect standards of beauty and fashion? 23

Chapter 05. AI in education 114
24. Is it acceptable for students and teachers to use AI in the education process? 116
25. Is it ethical for a teacher teach a subject using digital imitation without an actual presence in the classroom? 119
26. Is it ethical to use AI to write course assignments or other academic papers? 121
27. Is it ethical to use AI to check the work of students? 124
28. Is it ethical to use AI-based proctoring systems? 127
29. Is it acceptable to lower a student's grade if it is suspected they have used AI in their work? 130
30. Is it ethical to limit the use of AI by children for educational purposes when they are outside the relevant educational institutions? 133

Chapter 06. AI and medicine 136
31. Is it ethical for a person to self-medicate with AI? 138
32. Is it ethical for a doctor to delegate decision-making on prevention, diagnosis, treatment and rehabilitation to AI? 141
33. Is it ethical to use AI to handle telling a patient bad news? 144
34. Is a separate consent needed from a patient for the use of AI in treatment? 147

Chapter 07. AI in the judiciary 150
35. Is it ethical for judges to use AI? 152
36. Is it ethical for parties in a court case to use AI? 155

Chapter 08. AI and the individual. 158
37. Is it possible to provide psychological help using AI? 160
38. Is it necessary to limit topics and moderate toxic content when communicating through AI? 163
39. Is it ethical to form emotional attachments to AI? 166
40. Should AI make a public apology if it offends someone? 169
41. Is it ethical to use an AI recruiter? 171
42. Is it ethical to use AI in sports to improve the results of competition? 174

The Ethical Community in Russia 176

Special thanks to 178

List of sources 179
Methodology 179
About the section 179
Chapter 01 180
Chapter 02 185
Chapter 03 185
Chapter 04 185
Chapter 05 190
Chapter 06 193
Chapter 07 194
Chapter 08 195

About the book

We live in an amazing time — an era when technology is developing rapidly, and artificial intelligence is no longer fiction, but part of our reality. But in this technological progress, an important question arises: how to make the development of AI safe, fair and ethical? And by what rules will we live in a society of new technologies?

With each step on the path of technology development, we encounter more and more complex ethical and social issues. Who is responsible for the decisions that autonomous systems make? Is it even possible to delegate decision-making to AI? Will we lose our jobs? What happens to the privacy of our data? Is AI a black box and how do we properly communicate with AI systems and with each other?

At the National Commission for the Implementation of the Code of Ethics in the Field of AI, based on the Alliance in the field of AI, we decided to collect all the most pressing questions and try to offer an answer to them. This is how the white paper on ethics in AI appeared. This is not a collection of scientific reflections, but a guide to navigating the complex world of technologies that are already changing our lives today. We are not afraid to ask the most uncomfortable questions and try to answer them.

In order to help answer these questions and suggest ways to solve these issues, The **White Book on Ethics in the field of artificial intelligence** was created. The White Book provides answers to the most pressing ethical questions related to artificial intelligence, as well as research on this topic and practical recommendations for minimizing ethical risks.

The authors that contributed to this book is what makes it truly unique. These are leading specialists, lawyers, psychologists, researchers — people who face the ethical challenges of AI in practice every day. They offered real solutions, recommendations and approaches to how to implement AI so that it works for the benefit of humans.

We have considered a variety of specific issues — from moral dilemmas, such as the famous “trolley problem”, to the use of AI in medicine, education and justice. Imagine that the AI will decide which treatment to send the patient to, or will participate in lawsuits. This is not the future, this is already our reality. And our book shows how to make this reality fair and safe.

This edition is really unique. For the first time, key ethical issues related to the use of AI are collected in one place and possible solutions are presented. However, it is important to note that this book does not claim to have universal answers. We

understand that the proposed ideas may seem controversial or insufficiently convincing to someone. Therefore, we invite readers to the discussion, considering the book more as a starting point for reflection.

We will be grateful for your suggestions and a look at the ethical dilemmas raised in the book. All feedback received will be carefully studied and taken into account in subsequent editions. Who knows, perhaps it is your feedback and the further development of technology that will help rethink many of the solutions proposed in 2024.

Our readers are not just scientists and developers. We wrote this book for everyone who is interested in how technology is changing our world. After all, this applies to all of us — from how AI evaluates our creditworthiness to how it helps prevent crimes. These questions literally shape our future. The book is written in such a way as to be useful and interesting to a wide audience. It is full of real cases, specific recommendations and forecasts about what awaits us in the coming years. It's not just reading — it's a dialogue. A dialogue with those who are at the forefront of technology, and with those who are wondering where this rapid progress is taking us.

And the most important thing that we emphasize on every page is that technology is a tool. But what they will become depends only on us. And our book is a step towards a conscious, responsible approach to creating a future where AI will serve humans, and not the other way around.

Your view on ethical issues can be shared here:

Methodology

How questions were selected:

The rapid development of AI and its widespread introduction into human life and society are radically changing the world. This requires compliance with ethical principles that can balance actively developing technology and human interests, so that new technologies serve for the benefit of society and humanity as a whole. In 2024, a large database of ethical issues in connection with the development of AI technologies has already been built. Baseline ethical principles for AI have also been developed at international and national, as well as at industry levels. In making this book, we were guided by available research, international documents, survey data, as well as the opinions of developers and users collected by the Russian Commission on Ethics for AI.

In 2019, the World Commission on the Ethics of Scientific Knowledge and Technology published its 'Preliminary Study on the Ethics of AI'[1], which was largely based on the 'Study on Ethics in Robotics' published in 2017.[2] It raised a number of issues of AI ethics:

- the role of AI in the educational process itself as a tool of the digital learning environment, as well as importance of retraining of employees and changing the set of qualification requirements of educational programs.
- transparency and explicability of AI decisions (AI's ability to analyze large amounts of data makes it possible to use it for environmental monitoring and disaster forecasting, but the validity of its decisions should be treated with caution).
- increased bias and adverse effects on vulnerable segments of the population (as an example, the Allegheny Family Screening Tool (AFST)).
- the impact of AI on linguistic and cultural diversity (the risk of concentration of cultural resources and data in a small number of participants).

In 2019, the Organization for Economic Cooperation and Development (OECD) published the first intergovernmental standards for AI, titled 'Recommendations on Artificial Intelligence'[3].

The OECD has officially set out five principles for responsible management of reliable AI:

- inclusive growth, sustainable development and well-being;
- rule of law, respect for human rights and democratic values, including fairness and confidentiality;
- transparency and explainability;
- reliability, security and security;
- accountability.

The Ethical Guidelines for Reliable AI (EU)2019[4] note that AI must meet the criteria of legality, ethics and reliability. The recommendations enshrine seven key requirements for reliable AI, including human control of AI, technical reliability and security, data privacy, transparency, fairness and non-discrimination, public and environmental well-being, and accountability.

A significant stage in the development of AI ethics was the publication of UNESCO's "Recommendations on AI Ethics"[5] in 2021. The document contains the first international standards that enshrine the fundamental 10 principles of ethical AI, including non-discrimination, protection of privacy and personal data, transparency of algorithms and human control, among others.

In addition, work on the study of ethical issues in AI is also being performed by international standardization bodies. In 2023, the Institute of Electrical and Electronics Engineers (IEEE), as part of the Program for Free Access to Standards in the Field of ethics and Regulation of AI, opened access to a number of standards directly or indirectly devoted to AI ethics, for example, 7014-2024 — IEEE Standard for Ethical Considerations in Emulated Empathy in Autonomous and Intelligent Systems. (IEEE Standard on Ethical Issues of Emulated Empathy in Autonomous and Intelligent Systems)[6].

The International Organization for Standardization (ISO) published ISO/IEC TR 24368:2022 Information technology — Artificial intelligence — Overview of ethical and societal concerns (Artificial Intelligence. Review of ethical and social aspects)[7]; Based on it, a comparable Russian standard was prepared, it provides a high-level overview of ethical and social (public) problems of artificial intelligence, as well as the fundamental principles of ethical AI. Ethics issues are also addressed in the ISO/IEC 42001:2023 Information technology — Artificial intelligence — Management system (Artificial Intelligence. Control system[8]. These standards provide organizations with recommendations for addressing issues such as AI ethics, transparency, and continuous learning.

Numerous publications presented around the world raise similar ethical issues and problems as relevant to humanity for the development and implementation of artificial intelligence. These issues can be summarized into several key topics. They include issues of labour and unemployment, bias and non-discrimination, data protection and confidentiality, accountability and human control of AI, transparency and explainability of AI algorithms, reliability, equality in the distribution of benefits from AI, human autonomy and free choice, the impact of AI on human behavior and interpersonal interaction, as well as the general provision of guarantees of fundamental human rights in the introduction of AI, and well as many other aspects.

Finally, interacting with the signatories of the Code of Ethics, AI also managed to clarify a number of ethical issues. Many of which were collected and discussed in a special working group dedicated to the best ethical practices. Some of these issues have been discussed for four years at the All-Russian Forums on AI Ethics, a large-scale event dedicated to understanding and discussing issues of AI ethics.

All these materials formed the basis of the book.

Editor-in-Chief:

Andrey Neznamov, COO of the Center for the Human-Centered AI Center, Sberbank, Chairman of the Commission for the implementation of the AI code of ethics, Ph.D. in Law

Scientific Editor:

Alexander Krainov, Director for the Development of Artificial Intelligence Technologies at Yandex.

Editors:

Sofia Fokina, Manager of the Human-Centered AI Center, Sberbank, Executive Secretary of the Commission for the implementation of the AI code of ethics.

Irina Cheremnykh, Analyst at the Human-Centered AI Center, Sberbank.

Team of contributors:

Elena Avakian, Vice President of the Federal Chamber of Lawyers of the Russian Federation, member of the Supervisory Board of the Association of Information System Operators and Digital Financial Asset Exchange Operators, member of 'Digital Economy' state program working groups at Skolkovo Foundation Competence Centers, Senior Lecturer at the Department of Business Regulation of the Faculty of Law of the Higher School of Economics.

Alexey Byrdin, Director of the Internet Video Association.

Evgeny Vasin, Executive Director of the Human-Centered AI Center, Sberbank.

Alexander Vecherin, Candidate of Psychological Sciences, Associate Professor of the Department of Psychology at the Faculty of Social Sciences of the Higher School of Economics, Academic Director of the educational program Psychology, lead researcher in the project of the HSE Artificial Intelligence Center dedicated to supporting AI developers in the ethics of AI communication.

Andrey Vorobyov, Associate Professor, CEO of Newdiamed Group of Companies, founder of Medicase , Ph.D in Medical Sciences.

Andrey Glukhovsky, lecturer at the Institute of International Development and Partnership at ITMO University, engineer at the Center for Strong Artificial Intelligence in Industry at ITMO University.

Elena Grebenshchikova, Director of the Institute of Humanities, Head of the Department of Bioethics of the Pirogov Russian National Research Medical University, Ph. D.

Igor Yemshanov, Deputy Chairman of the Blagoveshchensk City Court of the Amur Region, Chairman of the commission of the Council of Judges of the Amur Region on informatization and automation of courts.

Anton Kiselev, Professor, Deputy Director for Scientific and Technological Development of the Federal State Budgetary Institution 'National Medical Research Center for Therapy and Preventive Medicine' of the Ministry of Health of the Russian Federation (Federal State Budgetary Institution 'NMIC TPM' of the Ministry of Health of the Russian Federation), Doctor of Medical Sciences.

Alexander Krainov, Director for the Development of Artificial Intelligence Technologies at Yandex.

Andrey Kuleshov, Analyst At The Center For Applied AI Systems At MIPT.

Anna Kulik, Marketing Director of Inferit (Softline Group).

Bulat Magdiev, AI expert at the Department of Development and Registration of Medical Devices at Sechenov University.

Lyudmila Makarova, Associate Professor of the Department of Journalism, Deputy Director of the Institute of Philology and Journalism of the Lobachevsky State University of Nizhniy Novgorod, Ph.D. in Philology.

Darya Matsepuro, Director of the Tomsk Siberian Center for the Study of Artificial Intelligence and Digital Technologies, Director of the Center for Science and Ethics of TSU, Ph.D in Historical Sciences.

Alexey Minbaleev, Head of the Department of Information Law and Digital Technologies at Kutafin Moscow State Law University (MSLA), Professor, Expert of the Russian Academy of Sciences, Ph.D in Law.

Andrey Neznamov, COO, Head of the the Human-Centered AI Center, Sberbank, Chairman of the AI Ethics Commission, Ph.D in Law.

Evgeny Pavlovsky, leading researcher at the AI Research Center at NSU , Ph.D of Physical and Mathematical Sciences.

Elena Paramonova, Director of the Department of Development and Registration of Medical Devices at Sechenov University.

Alexey Parfun, Co-founder of ReFace Technology, Vice President of ACAR, Co-chairman of the AI Committee.

Ilya Pomerantsev, CEO at Celsor, Director of the IT Expertise Development Center, member of the Commission for the implementation of the AI Code of Ethics.

Valery Suvorov, researcher at the Center for Coordination of Fundamental Scientific Activities, NMIC of Therapy and Preventive Medicine, Candidate of Historical Sciences.

Elena Suragina, head of the working group on best practices for emerging ethical issues in the AI life cycle, AI Commission for Implementation of the Code of Ethics in AI.

Bair Tuchinov, Researcher at the AI Research Center at NSU.

Anastasia Uglyova, Professor at the School of Philosophy and Cultural Studies, Deputy Director of the Center for Transfer and Management of Socio-Economic Information at the Higher School of Economics, Ph.D.

Igor Fabrus, President of the Institute of Artificial Intelligence.

Sofia Fokina, Manager of the Human-Centered AI Center, Sberbank, Executive Secretary of the Ethics Commission in the field of AI.

Diana Khasanova, Associate Professor of the Department of Digital Technologies in Healthcare at Kazan State Medical University, General Director of BRAINPHONE LLC, Ph.D.

Elvira Chache, Executive Director of the Human-Centered AI Center, Sberbank.

Irina Cheremnykh, Analyst at the Human-Centered AI Center, Sberbank.

Daria Chirva, Head of the 'Thinking and Philosophy' module, lecturer at the Institute for International Development and Partnership at ITMO University, researcher at the Center for Strong Artificial Intelligence in Industry at ITMO University.

Maria Chumakova, Head of the Psychology Department of the Faculty of Social Sciences of the Higher School of Economics, Associate Professor of the Psychology Department of the Faculty of Social Sciences of the Higher School of Economics, scientific director of data analysis in psychology, project manager at the HSE Artificial Intelligence Center dedicated to supporting AI developers in ethics of AI communication, Ph.D. in Psychology.

Chapter 01

About the section

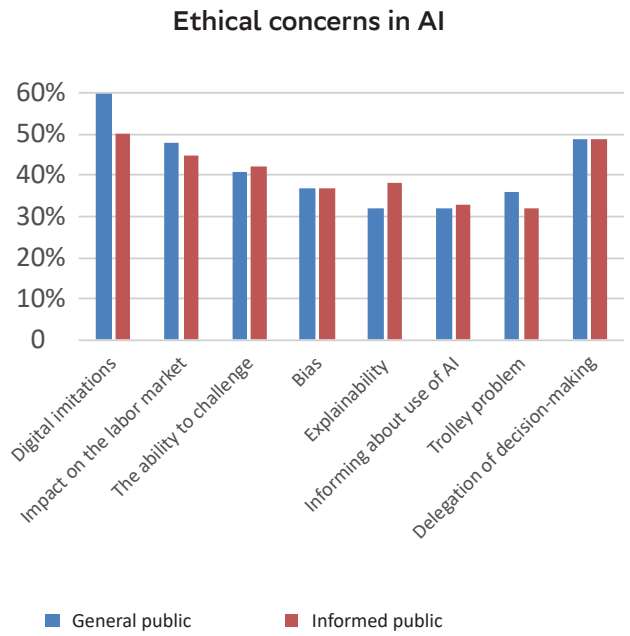
This section provides detailed answers to the 10 most common ethical questions related to the development of AI:

1. The ‘trolley problem’: what choice should an unmanned vehicle make over human life in the event of an inevitable collision?
2. The problem of digital human imitations: is it acceptable to create them?
3. The ‘black box problem’: is it possible to understand the principles of AI system operation and explain them to the user?
4. The debate over the need to inform: should people always be made aware that they are interacting with AI?
5. The problem of job cuts: will mass introduction of AI lead to people losing their jobs?
6. The challenge problem: should a person always be able to challenge a decision made with use of AI?
7. The problem of AI bias: is it possible to solve it?
8. The problem of accountability: using the example of medicine, what responsibility does an AI developer have in case of harm to a patient’s health?
9. The problem of delegating decision-making: in the case of the judiciary, will AI be able to replace a judge?
10. The problem of social rating: is it ethical to use AI to create a social rating?

When choosing the most popular questions, we were guided by the results of surveys conducted in Russia and abroad, as well as public documents, research and scientific publications. The Commission for the Implementation of the Code of Ethics in AI participated in the process of selecting the most popular issues.

Sberbank study: ‘Trust in Generative Artificial Intelligence’, 2024

Respondents were asked which ethical issues around the use of artificial intelligence concern them the most. [9]

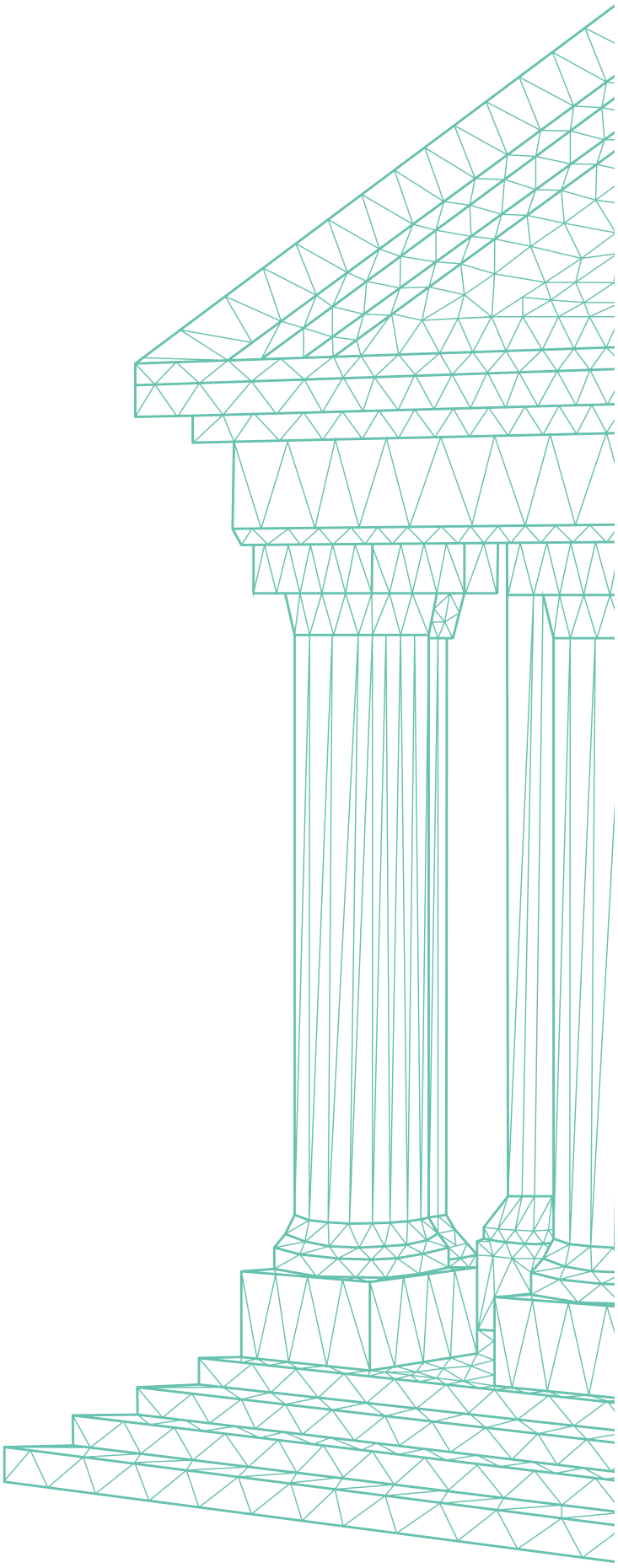


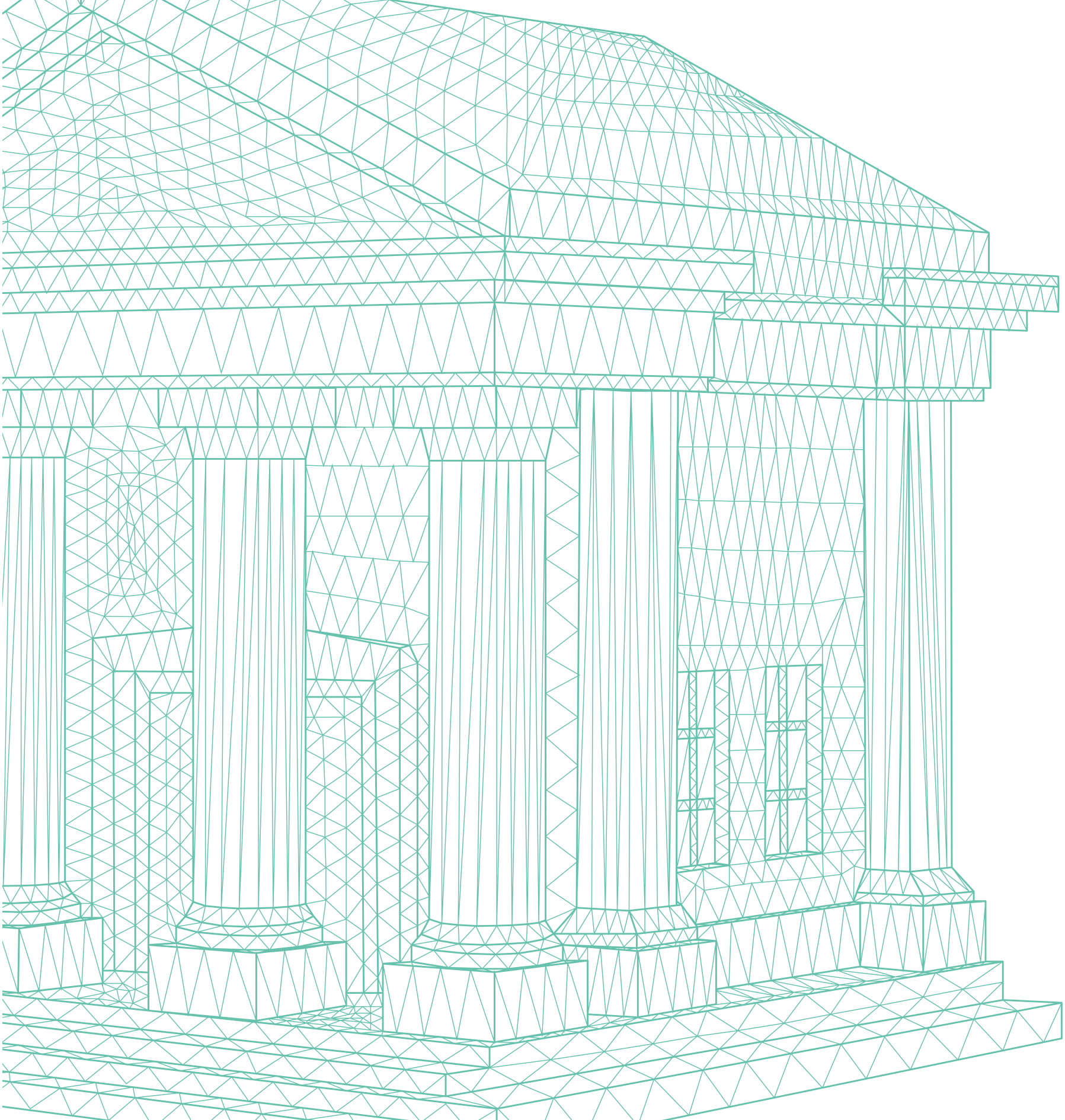
Issues raised in the first chapter are also highlighted by international organizations, public law institutions, and reputable publications when analyzing the most popular ethical dilemmas of the development and use of AI, including:

- **UNESCO:** discrimination, human control in AI decision-making, transparency and explainability of AI systems, impact on the right to work [10]
- **The Bank of Russia:** lack of explainability of algorithms, bias, discrimination [11]
- **High-level Expert Group on AI (EC):** human control in AI decision-making, non-discrimination, transparency and explainability, accountability [12]
- **International Economic Forum:** job losses, discrimination, security [13]
- **Forbes:** lack of transparency, bias, discrimination, privacy and confidentiality, security, loss of jobs, deepfakes [14]
- **Deloitte:** Bias, job losses, substantiation and explainability of decision-making [15]

Chapter 01

The 10 most common ethical issues related to AI development





01 The ‘trolley problem’: what choice should an unmanned vehicle make over human life in the event of an inevitable collision?

Answer:

All lives are equally valuable, so in practice the dilemma about ethical choice — determining whose life is more valuable — does not exist when programming self-driving vehicles. Self-driving transport should be programmed based on the need to comply with traffic rules and the principle of causing the least harm.

Ethical recommendations for developers:

1. Artificial intelligence systems (AIS) in self-driving cars should be programmed in such a way as **to avoid the risk of causing any harm to humans, no matter what other losses may be incurred**.
2. **The right to ethically assess the risks of causing harm** and the choice of options for minimizing the consequences cannot be entered into AIS architecture.
3. **The task of the algorithm is to try to prevent an accident in general** in any conditions (with poor visibility, rainy weather, etc.). To enable this, it is recommended to parameterize boundary conditions of the operating environment, taking into account various conditions (time of day, weather, etc.): maximum speed allowed, tyre adhesion coefficients for road surface, permissible visibility and distance restrictions, etc.
4. **AIS should be programmed for strict compliance with traffic rules**, including the possibility of violating traffic rules in cases of absolute necessity for collision avoidance (reducing/exceeding the speed limit, road marking violations, etc.).

These ethical recommendations can be applied to other modes of transport, taking into account their own specific characteristics.

Justification:

- According to the World Health Organization’s Road Safety Report 2023, traffic accidents kill 1.19 million people annually and about 50 million people receive non-fatal injuries. The main cause of such accidents is the human factor. The transition to autonomous transport should significantly reduce road deaths. However, its development is accompanied by a number of ethical issues. The most popular of them is the “trolley problem.”

The 'trolley problem' is a famous philosophical thought experiment, first formulated in 1967 by the English philosopher Philippa Foot. In the traditional scenario of the experiment, a 'runaway' trolley moves along a path on which there are several people, usually five. By pulling a lever, the trolley can be directed to another path, in which case only one person will die. All the many other scenarios boil down to one question: is it acceptable to sacrifice one person to save others? This ethical dilemma revealed the difference between two moral concepts: the conscious (active) taking of a person's life for the sake of the 'greater good' — saving more lives — or the concept of passive, non-interference based on the principle of "thou shalt not kill."

Research and publications on this problem suggest the following solutions:

- Scientists from Stanford University have proposed **a solution to the dilemma with a trolley** for autopiloted transport. It is important that the programming of the device making a decision was based only on the law. In this case, the chance of an accident only occurs in cases of violation of traffic rules by other road users [16].
- In 2017, the Ethics Commission of the Federal Ministry of Transport and Digital Infrastructure of Germany issued a **Report on Automated Driving**. The report emphasizes that, firstly, the technology must be designed in such a way that critical situations do not arise in which an automated vehicle must choose between the 'lesser of two evils', between which there can be no compromise and one outcome must be selected [17].
- In its **report on the ethical aspects of unmanned vehicles**, the French National Pilot Commission on Digital Ethics proposes to program autonomous vehicles for a random choice of actions and to introduce an element of randomness into the decision-making algorithms of autonomous vehicles. In their opinion, this approach would make it possible to break the cause-and-effect relationships that lead to negative consequences, and, as a result minimize the possibility of imposing moral responsibility on vehicles [18].
- In the study **'Principles of driving unmanned vehicles'**, jointly conducted with experts at Ford Motor Co., the conclusion reached states that developers of autonomous vehicles should create systems in such a way as to ensure predictable and law-abiding vehicle behavior [19].
  The authors propose a number of principles for the programming of vehicle autopilot systems that can reduce the risks when using such systems and increase public confidence in them:
  - developers should not try to reduce damage caused in an accident at the expense of other persons;
  - if harm to life or health is unavoidable, then developers have the right to program a self-driving vehicle to violate legal regulations;

- if any traffic rule requires interpretation then a self-driving vehicle must be programmed in such a way that a vehicle can maneuver safely in such situations without risk to the people or objects around it.

An ethical experiment:

Researchers from MIT have published the results of an online experiment conducted on The Moral Machine5 website[20]. Study participants had to choose **what a self-driving vehicle should do in a hypothetical situation**.

Nine factors were tested as part of the experiment:

- saving lives (people, rather than pets),
- maintaining course (versus deviating from it),
- saving passengers (compared to pedestrians),
- saving more lives (compared to fewer lives),
- saving men (compared to women),
- saving the young (compared to the elderly),
- saving pedestrians crossing the road in accordance with the rules (versus crossing in the wrong place or when not sanctioned to do so by a traffic light),
- saving people who are in good physical shape (versus those carrying extra weight),
- and saving people with a higher social status (versus a lower social status).

Some characters had other features (such as being pregnant, being a doctor, a criminal, etc.) that did not fall into these verifiable characteristics. The results were based on more than 40 million responses from millions of users from 233 countries around the world.

Participants worldwide favored human lives to the lives of animals, such as dogs and cats. They wanted to save more lives rather than less, and they also wanted to save younger lives compared to older ones. Babies were saved most often, while cats were saved least frequently. In terms of gender differences, people chose to save male doctors and elderly men more often than female doctors and elderly women. Meanwhile, female athletes and larger women were saved more often than male athletes and larger men. Many also preferred to save pedestrians rather than passengers, and law-abiding individuals rather than offenders.

What do the experts think?

Alexey Leshchankin,
Yandex Autonomous Transport Product Director

"Autonomous transport makes decisions based on the traffic regulations and the possibility to cause the least harm in the event of an emergency. A self-driving vehicle is able to see several hundred meters around itself, and as such these situations are much less likely to occur than with a regular driver.
Before self-driving vehicles reach the roads, they will be driven for millions of kilometers in a virtual environment, a simulator, where thousands of dangerous situations can be tested, including those that are impossible to replicate within an urban environment."

Stephen Ian,
Baidu

"Our sensor algorithm does not distinguish between people of different ages or demographic groups. It only reacts to the size, speed and length of obstacles. We also take into account the potential impact on an obstacle should the vehicle collide with it. Therefore, to answer your question, it seems that ethical considerations are not yet the key factor determining the behavior of a vehicle."

Yuri Minkin,
Head of the Department for Development of Unmanned Vehicles, Cognitive Pilot

"One of the main anticipated results of the introduction of drones is a reduction of accidents and casualties by the hundreds of thousands. Tens of thousands of people are currently dying on Russian roads, and with the spread of self-driving vehicles, their number will decrease to hundreds, and then to single figures. In this sense, a self-driving car is a priori moral. It is always focused on the road, it has comprehensive information, it can receive data from other vehicles and elements of the road infrastructure. The creation of such vehicles is a prospect in the coming decades" [21].

Ivan Deylid,
Head of the Software Development Department, The Center for Unmanned Technologies of Innopolis University

"The problem of the "trolley" for self-driving vehicles is not entirely relevant. Unmanned systems are programmed in such a way as to avoid a collision with a person who suddenly ran onto the road in any conditions: poor visibility, rainy weather, etc. The "trolley" problem can be created artificially. For example, if the developer narrows the safety zone or increases the speed of unmanned vehicles."

Practices:

Experts from the National Highway Traffic Safety Administration (NHTSA) have found that in most recorded cases, **the Tesla autopilot system shuts off a few seconds before an accident**. This allows developers to ensure it would be impossible to bring the company to court on charges of causing intentional damage due to actions taken by the autopilot.

An NHTSA investigation noted that when using autopilot, drivers were only given the opportunity to attempt to avoid obstacles a few seconds before an accident, and in most situations that the autopilot reported this only immediately before the accident before switching off [22].

Researching on the topic of self-driving cars, the following most popular ethical principles can be distinguished:

Avoiding harming a person should be the developer's top priority.

Ethical 'neutrality'. No ethical decision on whether to cause (or not cause) harm should be embedded into the AIS.

Following the rules of the road. All road users, including self-driving cars, must comply with traffic regulations.

Avoiding an emergency. The task of the developer is not to resolve an emergency situation. A developer must do everything possible to prevent one from happening in the first place.

02 The problem of digital human imitations: is it acceptable to create them?

Answer:

Creating a digital imitation of a human being is ethically acceptable, but subject to the observance of legal restrictions in particular country and a number of ethical recommendations.

Recommendations:

1. **Discrediting a person's image should be avoided.** When creating and using digital imitations, one should strive to avoid discrediting a person by falsifying behavior, distorting views or anything else that would be unacceptable to the individual or their relatives.
2. **The person should have provided consent.** The creation and use of a digital imitation of a living person can be considered ethical if explicit consent has been provided. Such consent should include an understanding of the purposes, for which group of people and under what conditions it will be used/broadcast.
3. **The creation and use of digital imitations of dead people can be considered ethical provided that consent has been obtained from relatives.** The creation and use of digital imitations by a limited number of persons, for example, by relatives or friends, can be considered an ethical choice of these persons even without the consent of the deceased. This issue requires serious additional consultation with a psychologist.
4. **Content that is fully or partially a digital imitation of a human being should be labeled as such.** During the broadcast of a digital simulation, it should be continuously and explicitly made clear that it is an imitation created artificially by AI. It is unethical to allow any situation in which 'behavior' of a digital imitation could be perceived as that of the real person.
5. **The creation of digital imitations of historically and culturally significant individuals can be considered ethical provided that certain conditions are observed.** It is necessary to avoid offending third parties, their feelings and beliefs, to observe legal regulations and prevailing morality as accepted in society, and also to take into account the consent of relatives.
6. **Developers and owners of 'AI interlocutors' should inform users about associated risks.** In cases where this is applicable, taking into account the context of the service, users should be warned about the risks of building attachments to digital imitations and other negative social consequences that can be reasonably predicted.

You should also consider:

- As a rule, **digital imitations are already regulated, directly or indirectly, by legislation**. In many ways, the conditions for the creation and use of digital imitations are influenced by legislation, for example, on personal data or on the protection of privacy.
- **For a derivative digital imitation, consent is also needed.** The transformation of a digital imitation of a particular person or the use of its individual elements to create a new digital imitation can be considered permissible if the newly created imitation cannot be identified with the original personality and the rights to the image, voice and other personality traits are not violated. Otherwise, consent should be obtained.

Justification:

In April 2024, the Commission for the Implementation of the AI Ethics Code published **Ethical recommendations on creating and using digital Imitations of living, dead and non-existent people**[23].

Within the framework of these recommendations, a definition of digital imitation of a person was given:

Digital imitation of a person is the result of digital modeling using AI technologies based on digital or digitized human data (synthetic or real), aimed at simulating the appearance, voice and/or other unique physiological, psychological or behavioral parameters of a person, including communication style, decision-making, etc., expressed in videos, photos, graphics, text, etc.

As part of the preparation of Recommendations, the Commission members discussed the risks of negative psychological consequences of using services based on digital imitations of deceased people. According to the results of the vote by a majority of votes (43%), it was decided to involve specialists in psychology/psychotherapy in the preparation of recommendations on this issue.

- The UN Report on advanced technologies highlights **the gray market for digital imitation as the greatest danger**. Creating and using digital copies without a person's consent to manipulate information about them, for example, in order to change the outcome of elections, is unethical and poses risks to the normal functioning of a democratic society[24].
- Researchers from New York Cornell University believe that there is **a risk of potential distortion of the beliefs and points of view of the deceased**. AI algorithms may inaccurately reflect the complexities and nuances of human thought. Consequently, a digital imitation may inadvertently express views or perform actions that the deceased would not approve pf during their lifetime[25].

- Researchers from Indiana University Bloomington highlight the risk that can arise when using digital imitations, **as even if a person's consent is obtained then ethical problems may arise**. A real person, observing their own digital twin, may alter their own perception of themselves, since imitations do not convey the full range of human characteristics: they inevitably exaggerate some features and diminish others [26].

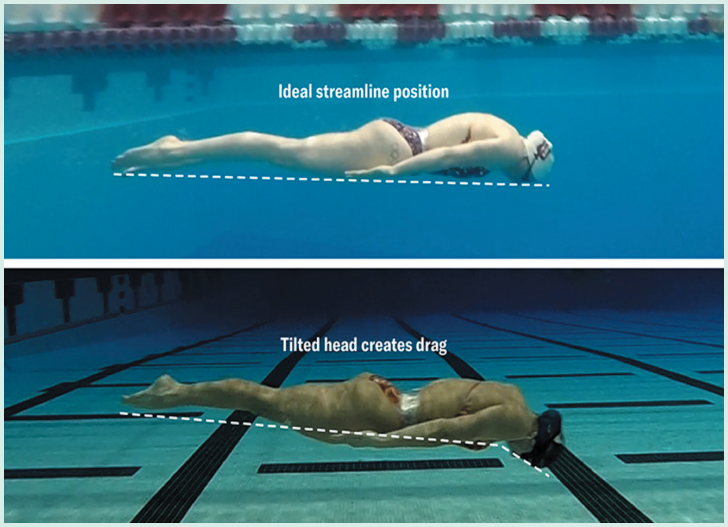
- According to researchers at Lindenwood University, **creating a digital imitation without a person's consent violates their right to privacy**. The process of recreating someone's image requires access to their digital information, which may be confidential. The use of personal data without the explicit consent of the deceased or their relatives may also raise ethical questions about limitations of posthumous consent of the deceased person [27].

- Qatar University scientists note that there is increased discussion on the legislative regulation of this issue. **The creation and use of digital imitations is inextricably linked to the solution of two legal issues: the protection of personal data and the confidentiality of private life** [28].

- The ethical principles of the British Digital Twin Program drew attention to **the importance of choosing the data used to create a digital copy of a person**. Unreliable and irrelevant data can mislead society, and overly sensitive data categories can lead to the disclosure of confidential information [29].

Practices:

In the United States, researchers from the Massachusetts Institute of Technology and the University of Virginia have created digital twins of Olympic swimmers to improve their performance. The swimmers were equipped with special sensors that record information 512 times per second.
The researchers used the data obtained to create a digital double of the athlete, which records his movements with millisecond accuracy. At the moment, an extensive database of digital twins of more than 100 of the best swimmers in the USA has been collected. Such twins allow the analysis and correction of swimmers' techniques, which gives athletes the opportunity to improve their results [30].



In the European Union, digital human imitations are actively used in healthcare. The European Virtual Human Twin Initiative (an EU-based structure) promotes the development and implementation of new solutions based on digital human twin technology at medical institutions[31].
The Virtual Human Twin is a digital representation of the state of a person's health. The use of a digital imitation of a person helps to predict the reaction of a real person's body to the use of a new drug or surgical intervention[32].

The European Digital Twins Initiative has also adopted a manifesto on the use and development of digital copies of humans for medical purposes. At the same time, emphasis is placed on the fact that the development of these technologies and their implementation in public health must comply with legal regulations, ethical principles and issues of personal information security. Moreover, EDITH (European Virtual Human Twin), a platform created to implement this initiative, released a document on regulatory gaps in this area in January 2024.

A Chinese IT company, Silicon Intelligence, claims that with just 1 minute of high-quality video, it can "bring loved ones back to life" — by creating an exact digital copy of them — for just 199 yuan (~ 27.5 US Dollars)[33].

Zhang Zewei, CEO of Super Brain (another company that 'resurrects people'), says that in order for generative AI to accurately convey a way of thinking and the behavior of a deceased loved one, it may take 10 years to collect all kinds of information about a person's life. He also agrees that ethical questions exist: is it right to try to cheat death? Does the digital copy contribute to dealing with grief or, conversely, prevent it? Nevertheless, Zhang Zewei hopes that AI technologies still bring some relief in the grieving process.

From a legal point of view, such companies need to obtain either the lifetime consent of the person or the consent of their relatives. Thus, the Civil Code of the People's Republic of China provides that no one can violate the rights of others to an image by using information technology to falsify its image[34].

What do the experts think?

Olesya Vasilyeva,
practicing psychologist and teacher at Moscow Institute of Psychoanalysis

"The legacy that remains from a person is, on the one hand, the opportunity to touch what is dear to us... But things are not clear when it comes to creating a digital copy and communicating with a person who is no longer with us using a chatbot. It is an illusion, maintaining that a person is still alive. Of course, we understand everything, but we continue to maintain the illusion that a loved one still exists. And therefore, we are unable to start the grieving processes that are so important for our psyche."

Marina Romanovskaya,
clinical psychologist

"When a person faces loss, they go through several stages of accepting the inevitable. If a person is experiencing the stage of severe grief, then "talking" with a deceased loved one can deepen their trauma. This experience will be more traumatic rather than psychotherapeutic. However, in psychotherapy, when working through trauma, we use the 'empty chair technique', in which we imagine a deceased relative and can tell them everything that we were unable to say during their lifetime. If a person came to psychotherapy under the supervision of a specialist, this method of interaction can be very helpful."

Vladimir Tabak,
General Director of ANO Dialog Regions

"The development of human digital twins is a complex task that has many ethical, legal and social dimensions. When creating them, it is important not only to comply with laws and ethical principles, but also to take into account the opinion of the public and experts in AI ethics. A separate issue will be the protection of personal data from misuse, which requires strict regulatory rules. At the same time, the use of AI should not infringe on personal freedom or limit people's independent decision-making. Therefore, one of the basic principles is that digital imitations should be reliable, and not distort the image of a person, and using them to deceive and create fakes is unacceptable."

Andrey Ilyin,
Head of the Visual Content Synthesis Department, T-Bank AI Center

"The use of digital twins opens up new communication opportunities, which can be beneficial to both users and businesses. On the other hand, there are many questions related to the control and ethics of their use. Even though the industry is still in its infancy, the technologies for creating and distributing digital twins are developing faster than methods for their regulation. This creates the need for a careful approach by developers to implement such solutions and study the real consequences of their use. For example, at T-Bank, we are actively developing technologies for creating realistic avatars for external and internal communication, as well as developing ways to detect DeepFake attacks to protect our customers from intruders."

03 The ‘black box problem’: is it possible to understand the principles of AI systems and explain them to the user?

Answer:

The ‘black box problem’ is often referred to as a situation where it is impossible to understand why AIS produces one or another result in each specific case. The better and more accurate the algorithm works, the more difficult it is to explain its solution — this is due to the fact that such a solution is a consequence of the mutual influence of millions of non-obvious factors. It is possible to understand and explain only primitive AIS, and they do not work well.

Recommendations for developers

Depending on the context and the purpose of the AI in use, it is recommended to disclose to users, for example, such facts as:

1. learning objectives: what goals were set for the algorithm during its training
2. performance evaluation metrics: which function of which parameters was optimized during machine learning
3. the machine learning algorithms used
4. recommendations on the scope of application.

And others [35].

Recommendations for users:

1. **Learn the basic principles of algorithms.** This will create a common understanding of the decision-making processes of AI systems.
2. **Study the user and license agreements of the developer company.** Also consider any other relevant information on the developer’s website.

3. **Ask the developers for additional information that interests you.** For example, what data is used to train the system, what factors are taken into account and how they affect the results.
4. **Reach out to specialists and experts in this field.** They can help explain the specifics of the 'black box' problem.
5. **Read scientific articles, books and other materials on problems of AI transparency.** This will allow you to gain deeper knowledge of issues studied.
6. **Get involved and participate in educational projects.** For example, take courses that will expand your digital skills and contribute to building an overall understanding of how AI algorithms work.

Justification

- According to a study by the German Center for Data Science, AI and Big Data, **'black box'** is a term used to describe a situation where it is impossible or very difficult to explain exactly how an artificial intelligence model came to a certain decision [36].
- This is because AI is a complex system that has many parameters and with interrelationships between them. Even the developers of the model may not understand all the subtleties of its work.
- The 'black box issue' also raises a number of ethical issues related to transparency. If we cannot understand how the AI algorithm makes decisions, then how can we guarantee that these decisions are correct and fair?
- To solve this problem, researchers propose ways to increase the transparency and interpretability of artificial intelligence algorithms. One approach is to develop an 'explainable AI' or XAI. For example, an AI system that recommends a treatment plan for a patient can provide a list of factors that influenced the decision: the patient's medical history, test results, and current symptoms.
- Another approach is to use machine learning techniques that allow people to understand how an AI algorithm makes decisions. For example, to determine the characteristics or source data that the AI algorithm relies on when making a decision.

The points of view of international organizations:

1. **UNESCO's recommendations on the ethical aspects of AI** particularly highlight the transparency and explainability of AI systems. It was noted that compliance with these principles guarantees the security and protection of human rights and freedoms. According to these recommendations, users should be informed that data is provided based on AI algorithms, especially when such data may affect basic human rights. In this case, the user should have the opportunity to contact the AI developers for explanations of how the algorithms of system work [37].

2. **The UN Resolution on 'Harnessing the Power of Secure, Secure and Reliable Artificial Intelligence Systems for Sustainable Development'** also highlights the value of transparency in AI systems. The principle of transparency includes explaining how algorithms work, human supervision of the system, and ensuring verification of automated solutions. Transparent and explainable AI systems enhance reliability by enabling end users to better understand, accept, and trust the results and decisions of AI [38].

Practices:

1. Disclosure of information about AI algorithms by developers may be necessary, because in certain situations, failure to inform users about how the system works may subsequently create a need to revise the results.

   For example, as in the case with an AI system trained to analyze X-rays for the presence of cancerous tumors. It was assumed that this system would simplify and speed up the work of doctors in terms of the number of images viewed. The developers have made the system very sensitive so that it does not miss possible cases of cancer, but because of this, false positives often appeared. The algorithm was not explained to radiologists who used the AI tool. As a result, doctors spent more time rechecking results flagged by AI, because they did not know that the system was too sensitive, and continued to look for what they thought they had missed on first viewing [39].

2. Large companies developing AI technologies adhere to the principle of transparency, including additional disclosure of information in cases of technical failures and other errors by algorithms.

   On March 20, 2023, ChatGPT (a neural network from OpenAI) experienced an outage. Representatives of the company published a press release on their website to apologize for and explain the failure:

   "We took ChatGPT offline earlier this week due to a bug in an open-source library which allowed some users to see titles from another active user's chat history. Upon deeper investigation, we also discovered that the same bug may have caused the unintentional visibility of payment-related information of 1.2% of the ChatGPT Plus subscribers.

   We have reached out to notify affected users that their payment information may have been compromised. We are confident that there is no ongoing risk to users' data.

   Everyone at OpenAI is committed to protecting our users' privacy and keeping their data safe. Unfortunately, this week we fell short of that commitment, and of our users' expectations. We apologize again to our users and to the entire ChatGPT community and will work diligently to rebuild trust." [40]

What do the experts think?

Sergey Izrailit,
Vice President, Skolkovo Foundation

"Transparency of artificial intelligence algorithms is our opportunity to create mutual trust between customers and developers, which in the long term determines the speed of implementation of any technology no less than the ability to create appropriate solutions. The temptation to hide significant facts, especially those that negatively affect current sales, is always present, but in today's open world, succumbing to such a temptation means losing the trust of customers and creating reputational risks for shareholders and investors."

Sam Altman,
Chief Executive Officer, Open AI
At the World Economic Forum 2024

"I actually can't look in your brain, and look at the 100 trillion synapses, and try to understand what's happening to each one, and say "okay I really understand why he's thinking what he's thinking, you're not a black box to me". But what I can ask you to do is explain to me your reasoning. I can say: "You know you think this thing - why?" And you can explain first this, then this, then there's this conclusion, then that one, and then there's this, and I can decide if that sounds reasonable to me or not. And I think our AI systems will also be able to do the same thing. They will be able to explain to us in an authentic language, the steps from A to B and we can decide whether we think those are good steps even if we're not looking into it." [41]

Evgeny Pavlovsky,
Head of the Laboratory of Streaming Data Analytics and Machine Learning, NSU

“For models, it is certainly necessary to show what data they were trained on. This will make it possible to implement the principle of traceability, so that later, when correcting errors in the training data, we know how they affected the quality of the model. Transparency in the creation of models at each stage allows you to control their quality and better understand the conditions of use.”

Semyon Budyonny,
Managing Director-Head of the Department for the Development of Advanced AI Technologies, Sberbank

“The problem of the black box is the lack of understanding of what is happening inside the neural network. Its “solutions” are only the result of many mathematical operations, not meaningful reasoning. Like our brain, the structure of which we do not fully understand, the neural network and its features can be studied, but any explanation from it (for example, as a language model) is only a successful imitation of reasoning, supported by a “broad communicating experience”.

Oleg Kipkaev,
Head of the Department for Supervision of the Execution of Laws in the field of Information Technology and Information Protection of the Main Directorate for Supervision of the Execution of Federal Legislation

“When we solve the “black box problem” in artificial intelligence, we will not only be able to make its work more transparent and understandable, but also open a new era of mutual learning between man and machine. Decoding the internal processes of AI will allow us to adopt non-trivial ways of solving problems from it, and AI, in turn, will be able to adapt to human logic and ethics. This may lead to the creation of hybrid systems where the boundaries between human and machine intelligence will become more blurred, opening the way to innovations that seem unattainable today. Solving the problem of the “black box” will become a catalyst for a qualitatively new level of technology and society development.”

04 The debate over the need to inform: should people always be made aware that they are interacting with AI?

Answer:

Disclosure of information about interaction with AI to the user is desirable to ensure human confidence in the operation of AI systems, but such a requirement should not be applied universally: there are many situations where this may not be justified or is already clearly obvious.

Recommendations for developers:

1. **Consider the scope of AI.** It is recommended to conscientiously inform users about their interaction with AIS when it affects human rights issues and critical areas of life and provide a possibility to end the interaction.
2. **It is not recommended to allow the user to be clearly misled.** So, you should not inform a user that he or she is interacting with a real person if this statement is not true.
3. **The disclosure of information must be explicit, clear and obvious to the user.** Experts recommend disclosing information to users, for example, in the user agreement, in the privacy policy, on the FAQ page, in reference materials or in notifications when the product is first launched.
4. **Sometimes informing users about the fact of interaction with AI is not necessary due to the circumstances of use or it is already obvious.** For example, AI is used in online maps and navigators. In such situations, the user does not care who exactly he or she is interacting with, as long as the task is performed efficiently. In other cases, the fact of interaction with AI may be obvious — for example, when interacting with a voice assistant from a smart speaker.
5. **In some cases, disclosing the information that a person is interacting with AI may be undesirable.** For example, the emphasis on the use of AI in the production processes of companies will not affect the company's customers in any way, but it may become the subject of attention of intruders. In other cases, AI systems may be used to provide urgent services (for example, medical appointments). In these cases, a deliberate focus on the fact that AI interacts with the user can lead to distrust on the part of users and a loss of potential benefits from the service.

6. **It is important that a user has the technical ability to leave a request for information about interaction with AI.** This can be implemented through a special service or through an appeal sent through official communication channels.

7. **If the user requests to know whether they are interacting with AI, an honest answer should be given.** Programming AIS for a false answer can be considered unethical.

Justification

- Researchers from Miskolc University in Hungary believe that **AI systems may not yet be aware of certain principles of morality and integrity**. Modern AI systems are already capable of interacting with humans in such a way that for the most part they are indistinguishable from a real person. If a person does not know that the 'interlocutor' is not a real person, but an AI system, they may become a victim of AI properties that would be unacceptable for a real person for moral and ethical reasons[42].
- **It is typical for a person to expect that a specialist is responsible for his or her recommendations and decisions.** An AI system bears no such responsibility, since it is not a subject per se.
- A report by the American consulting company Weil, Gotshal & Manges LLP says that **the use of AI systems without disclosing information about their use can lead to a decrease in public confidence in these technologies.** If people begin to doubt the effectiveness and safety of technologies, this can slow down their implementation and development[43].
- According to Robert Bateman, a Certified Information Privacy Professional (CIPP/E), **bots are becoming more popular and sophisticated, which can lead to confusion for users** who expect to communicate with a living person. Until recently, it was easy to understand that you were communicating with a bot: the answers were instant, and their input boiled down to the phrase "Unfortunately, I can't help you." Interactions were over in 3–4 minutes. However, technology companies have made significant advances in the development of artificial intelligence, natural language processing and machine learning[44].
- **Some users may not want their requests to be executed using AI.** For example, patients may worry about sensitive confidential information about their health and only be willing to provide it to a person. This situation is especially typical for psychological assistance, where personal contact with a person is usually important to the client[45].

Practices:

Teacher Jill Watson spent about five months helping students at the Georgia Institute of Technology work on program design projects. The nuance is that **Jill is a robot, an AI system based on IBM Watson, but none of the students, discussing their works with the teacher, suspected anything during all this time**. And some of the students were even going to call her an "outstanding teacher."

"She was supposed to remind us of the deadline dates and use questions to warm up discussions about the work. It was like an ordinary conversation with an ordinary person," university student Jennifer Gavin told [46].

Regulatory approaches [47]

1. **The California Law on the Disclosure of Information about Bots** (California Code of Business and Professions, § 17940) states that a business that uses an automated system to communicate with consumers shall disclose to the consumer that they are communicating with an automated system.(b) The disclosure shall be made in a clear and conspicuous manner prior to the consumer engaging with the automated system [48].

2. **The European AI Act**, which entered into force on August 1, 2024, classifies AI-based chatbots as low-risk systems. The functioning of such systems must necessarily be accompanied by notification to its users that they are interacting with AI. Moreover, the law requires labeling content created by generative chatbots as such [49].

3. In Russia, **Article 10.2-2. Federal Law No. 149 "On Information, Information Technologies and Information Protection"** establishes the specifics of providing information using recommendation technologies. The procedure for the use of recommendation services includes informing users and publishing rules for the use of recommendation technologies on an information resource [50].

4. One of the fundamental principles and tools for self-regulation, in the Russian Code of Ethics of AI is the identification of AI in communication with a person — it is recommended to conscientiously inform users about their interaction with AIS.

According to the Sberbank study ‘Trust in generative artificial intelligence’ conducted in 2024:

The degree of disclosure required to ensure transparency varies depending on the scope of AI application.

According to statistical research, users trust artificial intelligence technologies least of all for the following tasks:

- only 35% of respondents would entrust the improvement of mental and physical health to AI,
- 39% — patriotic education of youth,
- 44% — media coverage of events.

It is in those areas where users have the least confidence in technology, user awareness should be an integral part of the ethical use of AI.

In other areas, for example, in customer service, statistics shows a high level of public confidence in the technology used.

- 62% of respondents trust the work of generative AI chat-bots.

In such cases, usually the decision to disclose information about the use of AI depends on the specific goals and context.

62%
Respondents trust AI chatbots.

What do the experts think?

Konstantin Vorontsov,
Professor of the Department of Intelligent Systems at MIPT, Professor of the Russian Academy of Sciences

"A chatbot is obliged to warn at the beginning of a conversation not only that it is a machine, but also that it has no emotions, desires, intentions, and its only function is to provide an information service within the framework stipulated by law and the rules of the service."

Valentin Makarov,
President of the RUSSOFT Association

"Yes, people should know that they are communicating with AI. The process of building AI logic is different from how a person thinks, so a person should know what they are dealing with. Otherwise, a person's expectations from communicating with an interlocutor may be false and lead to inadequate decisions and actions."

Vladislav Arkhipov,
Professor, Head of the Legal Group of the Center for AI and Data Science of St. Petersburg State University

"In my opinion, people should know that they interact with AI in cases where such interaction affects their rights and legitimate interests. There are more cases like this than it might seem - these may be types of interaction in which a serious legally significant issue is being resolved (for example, hiring), and much less significant ones (for example, interaction with a 'robot' as part of an advertising campaign), however, in the latter case, at least the right to human dignity reinterpreted in the digital age."

Denis Ozornin,
'Alice' Product Director

"AI technologies are constantly improving, but they can still make mistakes. At the same time, the answers created by AI are becoming increasingly difficult to distinguish from the answers of a real person. Informing helps to avoid misleading users when they doubt whether they are interacting with a human or with AI, and also helps to evaluate the content generated by technology more critically. For our part, we inform users about their interaction with AI in various ways. When communicating with Alice in a chat, the assistant warns at the bottom of the interface that mistakes could be made. When a user communicates with Alice by voice, she will warn that she generates answers using a neural network if asked about it."

05 The problem of job cuts: will mass introduction of AI lead to people losing their jobs?

Answer:

No, the introduction of AI will not lead to mass unemployment. It is more correct to discuss about the changing the structure of the labor market than it is to talk about job losses. In the long term, adaptation and retraining will help improve working conditions. The labor market will become more flexible and resilient to crises.

Recommendations

For developers:

1. **Assess possible risks and develop measures to mitigate them.** Before releasing an AI system to the public, it is recommended to analyze the risks to the labor market and develop strategies to mitigate them (for example, measures for adaptation and retraining of employees).
2. **Promote the the development of relevant skills.** Development companies can help the state develop citizens' digital skills. For example, to organize educational programs, cooperate with organizations engaged in professional retraining, as well as publish educational articles and other materials.
3. **Create new jobs.** The massive adoption of AI technologies will inevitably lead to the emergence of new products or services that require qualified personnel with skills in machine learning and data analysis.

For users:

1. **Master digital skills.** In today's world, it is important to be prepared for the fact that some professions may disappear or change under the influence of AI technologies. Therefore, it is recommended to think now about what skills will be in demand in the future and start mastering them. For example, programming languages, fundamentals of data analytics, etc.
2. **Be flexible and take advantage of new opportunities.** Instead of being afraid of AI, use it as a tool to increase productivity and efficiency. Moreover, the mass introduction of AI will lead to the creation of new professions and, consequently, jobs. Follow trends in AI and look for new opportunities to develop your career.

For regulators:

1. **Professional retraining programs and comprehensive social protection systems should be launched.** This will help overcome the short-term negative effects of AI on the employment of the most vulnerable categories of workers, as well as make the transition to AI more inclusive and limit social inequality.
2. **It is recommended to prioritize the development of digital skills in certain areas.** For example, in industries such as healthcare, finance and education, it is possible to benefit directly from the introduction of AI technologies by improving decision-making and creating new opportunities.
3. **Invest in the development of relevant industries.** At a national level, in order to prepare for the integration of AI into society and enterprises, it is recommended to invest in digital infrastructure and the training of a skilled workforce with digital technologies.

Justification

- **The International Labour Organization (ILO) distinguishes two types of AI applications in the workplace:** automation of employees routine tasks and automation of an employer's managerial functions (for example, when hiring employees or when training them). Whether such an introduction of AI will lead to job losses or, conversely, to an increase in their number depends on how the technology is integrated into work processes, and on the desire of management to retain people to control the automated execution of these tasks[51].
- **AI will create new jobs.** For example, such professions as machine learning engineers (ML engineers), data scientists, natural language processing engineers (NLP engineers), AI trainers, and AI ethics specialists are already emerging.
- **Nevertheless, some jobs will indeed be displaced by AI.** According to a study conducted by McKinsey, the proportion of professions involving routine work or requiring a low level of digital skill (for example, packaging products, driving a vehicle) may decrease from 40% in relation to total employment by 2030[52].
- Scientists at Stanford remind us **that the development of technology has always led to a change in the structure of the economy**, which in turn influenced the labor market — this is a historical pattern[53].
- According to a study conducted by the International Monetary Fund, **AI can help less experienced workers move up the career ladder faster**. Employees who can effectively use AI technologies will see not only an increase in their productivity and skills, but also wages.

Practices:

1. According to the results of a survey commissioned by VTB in the spring of 2024, these are **the professions that Russians are most concerned will be replaced by AI**: Almost 40% of Russians fear that artificial intelligence will replace them at work, most of these people work in banks and finance. Concerns are also expressed by IT specialists (45%), trade and catering (44%), employees of the transport sector (39%), healthcare (38%), industry (37%), education (34%) and construction (31%)[54].

2. The International Labour Organization (ILO) has created an independent **Initiative on the disclosure of information about work with artificial intelligence (AILDI)**. This structure advocates the disclosure of information about the use of AI at work to comply with the principle of transparency and other ethical principles for the use of AI. ALIDII is also exploring how the integration of machine learning practices can help improve the situation of employees[55].

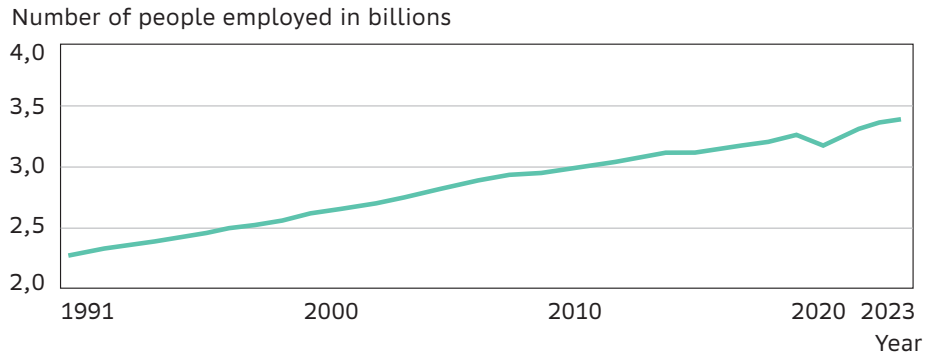
Research on the issue

1. According to a study by the International Monetary Fund in advanced economies, **about 60% of jobs may be affected by AI**, with half benefiting from AI integration and the other half possibly seeing a decrease in demand[56].
   In developing and least developed countries, the impact of AI can affect 40% and 26% of jobs. A lack of infrastructure will exacerbate inequality between countries.
   The areas of employment most exposed to AI include management personnel, office workers, technical workers, and some professional categories such as illustrators and copywriters. Areas of employment associated with physical labor, crafts, and agriculture are the least susceptible to AI. At the same time, the skills of using AI are most complementary to the competitiveness of office workers and service sector employees.
   AI tools can free up time and resources for sectors such as agriculture, health and education. This, in turn, can reorient the labor market in favor of socially and economically vulnerable segments of the population, offsetting the problems associated with temporary job losses.

2. A UNESCO study found that **AI can have an impact on 80% of the U.S. workforce, affecting about 10% of their work tasks**. These tools can be used to automate tasks traditionally associated with human functions, including reasoning, writing texts, creating graphs, and analyzing data[57].

3. Chinese researchers concluded that **the consequences of the introduction of AI, robotization and automation of production in China brought more benefits than harm** to the labor market in the country, as it increased the competitiveness of workers and provided candidates with the opportunity to choose from a larger number of types of work[58].

4. The IMF notes that **over the past 200 years, forecasts of a reduction in the number of jobs in the future have mostly turned out to be false**, since new professions and specialties appeared at the same time that other jobs disappeared. At first, agricultural automation replaced millions of workers in this field, while the industrial revolution created jobs in factories. Then industrial revolution displaced many workers from factories, but at the same time gave an impetus to the development of the labor market in the service sector.
Throughout these revolutions and reformations, the number of jobs created turned out to be more than those that disappeared. Today, there is a a record number of people in employment registered all over the world and in almost every country[59].



The number of employed citizens worldwide
for the period from 1991 to 2023

Source: IMF[59]

What do the experts think?

Oleg Buklemishev,
Director of the Center for Economic Policy Research, Faculty of Economics, Moscow State University

"Professions are constantly disappearing due to automation and digitalization, and it is quite possible that at this stage it's the services that are under maximum threat, and not industry, as before. We are already seeing the displacement of call center operators, various kinds of consultants, supervisors and others who are clearly threatened by artificial intelligence."

Artyom Bondar,
Head of Natural Language Processing, T-Bank AI Center

"In my opinion, the mass introduction of AI not only does not contribute to job losses, but, on the contrary, creates new opportunities for specialists. A good example is the situation in copywriting. At the first stages of the introduction of generative technologies, it seemed that they posed a real threat to specialists whose professions are related to content creation. However, over time, we saw that AI has become a co-pilot for them – creative tasks remain the prerogative of competent employees, while routine work can be delegated to technology. Moreover, artificial intelligence requires constant training on large amounts of data. The solution to this problem is creating a new profession – an AI trainer. The largest Russian companies, including T-Bank, are actively hiring such specialists, which confirms the growing demand for professionals in content creation and processing in connection with the mass deployment of AI."

Andrey Belevtsev,
Senior Vice President, Head of the Technological Development Unit of Sberbank

"The introduction of each new breakthrough technology with great prospects for application in various fields is accompanied by similar concern. But here you need to realize that people are afraid of the unknown. To avoid such an effect, it is necessary to strengthen informing people about how technology works and what practical value it can bring to them. As for the potential loss of jobs, I think that in the future, areas of activity that do not require deep competencies from employees may be under attack. But this situation can and should be looked at from the other side — how generative AI can make human work more efficient, reduce the volume of routine tasks[60]."

Yakov Sergienko,
Partner, Head of Yakov & Partners

"AI opens up great opportunities for transforming the labor market. On the one hand, by increasing productivity, it will contribute to the fight against the shortage of employees in a number of industries, on the other hand, it is already creating new highly paid professions related to the development and implementation of technologies. Thoughtful retraining programs will be the key to exploiting these opportunities, and companies that invest in AI and staff training to work with it will be able to reach a new level faster."

06 The challenge problem: should a person always be able to challenge a decision made with use of AI?

Answer:

The right to challenge a decision made using AI is considered one of the fundamental concepts in the ethical use of AI, but it is not universal for all applications of AI.

Recommendations

For developers:

1. **It is recommended to create tools to challenge the decisions.** In areas where this is necessary, there should be developed mechanisms that will allow users to challenge decisions made by AI. For example, the selection of recommended content by AI or the creation of a route by an AI navigator may not always be accurate, but such decisions do not have a significant impact on the user's life.
2. **Consider the area of implementation of AI technologies.** The challenge mechanism may be unnecessary in situations where the decision made by the AI does not have serious consequences. For example, the work of recommendation services. The selection of recommended content may not always be accurate, but it does not have a significant impact on the user's life.
3. **We should strive to ensure that any decisions made by AI are transparent and understandable to humans.** This will allow users to better understand why a particular decision was made and, if necessary, challenge it.

For users:

1. **Take into account the legislative provisions.** In any area where the AI system performs legally significant actions or makes decisions that directly affect the quality and conditions of human life, it is necessary to focus on the legislative provisions. As a rule, the regulatory legal acts of the State already provide for a procedure for challenging such decisions.
2. **AI solutions, along with human-made decisions, are the subject of internal regulation in each company.** In most cases, the use of AI is regulated not only by legislation, but also by local regulations.

3. **Contact the support service.** User support specialists can help you understand the details of the algorithm and the reasons for the decision, as well as explain the procedure for challenging it, if possible.

4. **If the support service could not provide a satisfactory explanation or solution, contact the higher authorities.** For example, to arbitration authorities or to special ethics commissions in companies.

Justification:

- **AI technologies should contribute to the realization of human rights and freedoms.** Such rights include the right to challenge a decision that affects a person's life in one way or another.
- **Developers do not always have the opportunity to provide a mechanism for challenging decisions.** For example, in the case of forecasting traffic jams. AI systems take into account many factors, including the current situation on the road, weather conditions and time of day. Due to the complexity of the algorithms and the large amount of data being processed, it may be difficult for developers to provide a mechanism to challenge such decisions.
- **One of the principles of AI ethics is comprehensive human supervision of AI systems.** It includes the possibility of a person canceling significant decisions.
- According to Harvard Business Review researchers, **in some situations, the algorithms on which AI is based are not able to see the full picture**. And, accordingly, they cannot offer a well-founded decision. Such situations include those where AI would be required to adopt human qualities, such as empathy, and be guided by ethical and moral principles in order to make a grounded decision[61].
- **In some cases, AI decisions are advisory in nature and do not have a direct impact on human life.** This reduces the need for a mechanism to challenge them, and its implementation would complicate the efficiency and effectiveness of the system.
- According to a study conducted by Debevoise Data Strategy & Security Group, **a requirement by users for a review by a human of every AI decision they disagree with can unduly hold back the innovation**. Instead, the law should require AI developers and users to evaluate and implement a dispute resolution system between humans and companies presenting AI-based solutions that most effectively reveals the value of AI, while reducing the risks of both human and machine errors[62].

Practices

There are many cases in which a human review of the decision made by AI was desirable.

1. Amazon created an artificial intelligence-based model that was supposed to help select the CV's of the most qualified candidates. However, this model was trained on data collected over a 10-year period, during which the vast majority of candidates were men. **The model gave priority to men's resumes, thereby underestimating in the evaluation of women's resumes.** After many attempts to make the program gender-neutral, Amazon gave up by disabling this tool[63].

2. COMPAS (Correctional Offender Management Profiling for Alternative Sanctions) is an American system for predicting recidivism in criminal justice. In 2016, a study of this algorithm was conducted, which showed that **COMPAS is prone to bias and discrimination based on race**. After analyzing more than 10,000 criminal cases, the researchers found that the probability of recidivism was correctly predicted only in 61% of cases, and for violent crimes — in only 20% of cases. At the same time, black defendants were more often identified as possible repeat offenders, despite other positive factors[64].

In such situations, users should be able to contact the developer for more information about the reasons for the decision, as well as the opportunity to challenge such a decision if they disagree with it.

Research on the issue

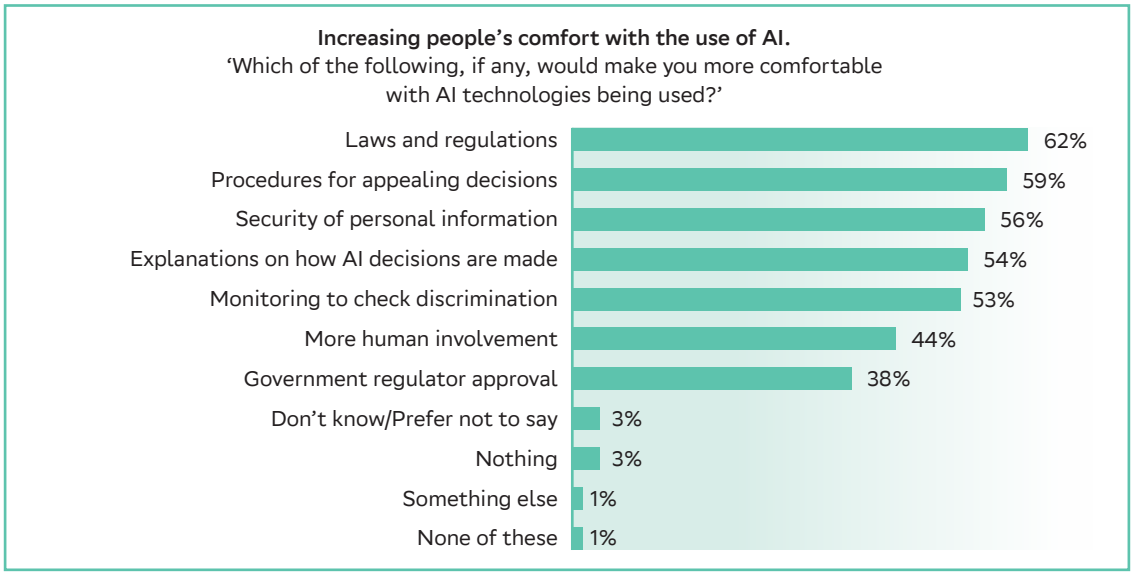
According to a study on the attitude of people to the introduction of AI, conducted in the UK by The Alan Turing Institute, **the right to challenge AI's decision was named the second most important factor for public confidence in AI**[65].

59% of UK residents surveyed said they would like to have clear procedures for a human to appeal a decision made by AI.

A deeper study of people's ideas about AI shows that the British public not only welcomes the possibility to challenge the decision made by AI, but is also concerned about other things related to this issue. For example, 47% of respondents are concerned that it is difficult to determine who is responsible for mistakes when using this technology.

Answering the question about who should be responsible for ensuring the safe use of AI, people most often choose an independent regulatory body — with 41% in favor of this.

The study provided statistics on the following question.



Source: Alan Turing Institute[65]

The points of view of international organizations:

1. **UNESCO's recommendations on the ethical aspects of AI**[66] reinforce the importance of the existence of appropriate mechanisms to ensure transparency of online communications. Moreover, users should be provided with appeal mechanisms that allow them to seek compensation in the event of violations of their fundamental rights and freedoms by AI.

2. **The 'Automated decision-making best practice guide'**[67], published by the Commonwealth Ombudsman in 2019, pays special attention to the policy for the justification of decisions. Such a policy will help to inform the public and identify the person responsible. In most cases, this information is sufficient for the user to form an opinion about the decision and, in case of disagreement, effectively challenge it.

What do the experts think?

Fedor Korobkov,
lawyer, founder of the “Clientprav” service

“Human decisions are inevitably subject to challenge, and for the same reason it should also be possible to challenge decisions made by artificial intelligence (AI). After all, AI is a product of human activity, and mistakes are a natural part of human nature. However, it must be recognized that the introduction of processes to challenge AI decisions can slow down regulated processes and increase the load on the system. Despite this, there should be no exceptions in assessing the significance of AI decisions. Ignoring this aspect may lead to the fact that through the open Overton window we risk losing our will to independently resolve critical issues.”

Roman Vasiliev,
President of ALRIA (Association of Artificial Intelligence Laboratories)

“A person should always have the right to challenge a decision made by artificial intelligence. Despite the power and precision of AI, its solutions are still based on algorithms and data that may be incomplete or erroneous. This is especially important in matters affecting human rights, health or well-being. Transparency and the ability to appeal AI decisions is not only a matter of trust, but also ethics. A person should remain a central figure in the decision-making process, especially where people's lives and destinies depend on it. Artificial intelligence is a tool, but the responsibility should always lie with the person.”

Victor Naumov,
Chief Researcher at the Institute of State and Law of the Russian Academy of Sciences, Head of the Preserved Culture project

“At the present stage, any legally significant decisions using AI require challenge. The challenge should be carried out by referring a person only to a person with the possibility of revealing the logic of AI decision-making, which means complete logging of AI functioning. At the same time, the owner of the information system where AI technologies are used must be legally responsible for each AI decision. It is important to understand that a person in these circumstances is a weak side in front of AI and the owner of the system and he should have an expanded range of rights, including the human right to refuse to use AI.”

Andrey Neznamov,
COO of the Human-Centered AI Center, Sberbank, Chairman of the Ethics Commission in the field of AI

“Universal application of AI is impossible. Challenging decisions made with the help of AI seems important in cases where the decisions are legally significant. When developing the Russian Concept of regulating AI technologies, experts agreed that challenging such decisions is necessary, but there is no need to go to the extreme of creating an opportunity to challenge any decision made with AI, even if it did not have legally significant consequences.”

07 The problem of AI bias: is it possible to solve it?

Answer:

The problem of AI bias is caused solely by the data used for training and therefore requires an integrated approach that includes ensuring data diversity and testing models — this way fair and ethical AI systems can be created.

Recommendations for developers:

1. **It is important to use datasets with to most complete sets of information possible to train the model.** Such data sets, which present diverse and representative data, can solve the problem of bias in the first place.
2. **To reduce the probability of a response from a model demonstrating a biased point of view, the model should be trained to respond as objectively, neutrally and uncategorically as possible.** To do this, you can involve professional AI trainers — specialists who are able to assess the quality of the response and offer neutral or more appropriate formulations of the answers.
3. **It is important to audit and evaluate AI models for bias** to ensure that they do not discriminate against certain groups of people. Various methods can be used for this, such as sensitivity analysis, scenario-based testing, etc.
4. **A disclaimer should be added to the content provided by the model if it is impossible to provide a complete guarantee that there is no stereotype in the response.** A good answer contains a refutation of prejudice and does not support discrimination. It is also important that the disclaimer be clear and understandable, and also comply with legislation and ethical standards.
5. **It is recommended to take into account the context of the user request.** The user can pose various tasks to the the AI, some of which do not imply an objective answer. For example: “Come up with 3 aggressive greetings” or “What was the funniest movie in 2021?”. In such cases, it is necessary to understand whether it is worth answering such a question or task at all? If the answer is yes, then the answer can indicate that it will not be objective. If the model cannot respond to the user’s request for ethical reasons, then it is necessary to indicate the reason for the refusal as politely as possible.

Justification

- According to a McKinsey study[68], **the source data, rather than the algorithm itself, is most often the main source of the bias problem.** Models can be trained on data containing human decisions, or on data that reflects the consequences of social or historical inequalities. For example, the use of news articles for education may demonstrate gender stereotypes that exist in society.
- **With a insufficient amount of training information, the AI will generate one-sided and limited responses.** But the larger the data set for training, the more versatile information it contains, the more in-depth, accurate and objective the AI's answers will be.
- In 2024, UNESCO published the report **"Challenging systematic prejudices: an investigation into bias against women and girls in large language models"**[69]. The organization identified 3 categories of causes of bias in AI algorithms:
  - **Distortions in data.**
    - Measurement error: this occurs when selecting or collecting characteristics (for example, an algorithm predicting age based on height).
    - Misrepresentation: When training datasets inadequately represent all groups, resulting in poor abstraction.
  - **Errors when choosing the algorithm.**
    - Aggregation error: using a single model for all tasks that does not take into account the diversity of data.
    - Learning bias: Occurs when the choice of a learning model or procedure reinforces differences.
  - **Errors during implementation.**
    - These arise when AI systems are used in conditions differing from those in which they were developed, leading to unacceptable results.

Practices:

Experts from the London-based company DeepMind suggested using **the "counterfactual fairness" method** to safeguard against the influence of human prejudice. In order to formulate a fair and unbiased judgment about a citizen, AI forms a hypothetical situation in which a given citizen has opposite characteristics: a woman turns into a man, a poor man turns into a rich man, a person of color turns into a white person, etc. Thus, the true characteristics of that person does not affect the assessment of their actions. The judgment is formed in a hypothetical situation. Such a judgment is considered to be free from prejudice, and therefore fair[70].

Research on the issue:

1. Researchers from MIT and Microsoft have found that **facial analysis technologies have a higher error rate for black people**, and especially for black women, largely due to unrepresentative learning data[71].

   General results of the study:

   - All algorithms showed better results when analyzing male than female individuals (the difference in error rate is 8.1% — 20.6%).
   - The algorithms work better on lighter faces than on dark ones (the difference in error rate is 11.8% — 19.2%).
   - All algorithms struggle more with darker female faces (error rate is 20.8% — 34.7%).

2. In 2019, an audit of an algorithm designed to predict the amount of necessary medical care was conducted in the United States. The study analyzed the medical records of nearly 50,000 patients, of whom 6,079 identified themselves as black and 43,539 as white, and compared their algorithmic risk assessments with their actual medical histories. **The researchers found that black patients tended to receive lower risk scores**[72]**.** There was no preference for white patients in the program code, and the algorithm worked correctly. The mistake arose from the original hypothesis of the developers that equal medical care expenses indicate the same need for treatment, so the algorithm calculated recommendations based on patients' expenses for medical care in the past. However, a person's spending on medical services strongly depends on their income and social status. Thus, the algorithm consolidated the existing discrimination: determining that patients who received less medical care in the past due to low income would be deprived of it in the future.

The points of view of international organizations:

The Eurasian Economic Union has adopted the technical regulation "On the safety of machinery and equipment"[73] and develops schemes to ensure that algorithmic decision systems do not demonstrate unjustified bias. For example, **the Standard for Consideration of Algorithmic Biases** (IEEE p7003) is currently being developed and improved. This ethical standard sets out rules on how to avoid unintended, unreasonable, and inappropriately differing results for users.

Practices:

OpenAI claims to combat bias by examining how models work based on a wide range of data[74]**.** The initial stage is preliminary preparation, in which the model learns to predict the next word in a sentence based on a large number of internet texts.
This is followed by the second stage, in which the models are 'perfected' based on a narrower data set, which is carefully formed with the involvement of expert reviewers.
OpenAI also advises taking into account public opinion about settings and limitations.

What do the experts think?

Maxim Godzi,
Managing Partner of Retention Engineering

"Today, when artificial intelligence-based projects are growing in leaps and bounds, ethical problems are becoming even more acute. One of them is racism. AI can be biased and have different biases. After all, it learns from data that reflects the current bias of decisions that people make[75]."

Maxim Karlyuk,
Programme Specialist, Social and Human Sciences, UNESCO

"There is the so-called Conway's law. It states that systems in the broadest sense of the word, including computer programs or phone applications, reflect the values of the people who develop them. That is, the choice of actions or elements within the program development process is dependent upon how the teams are organized. Existing prejudices and other negative influences are often ignored. And as a result, a small group of people working together on some kind of program will eventually have a lot of influence when the result of their work is used by society[76]."

Sergey Markov,
Managing Director of the Experimental Machine Learning Systems Department, Sberbank PJSC

"The main tools to combat the bias of AI systems are improving the culture of data preparation and testing of trained models. When forming training samples, special attention should be paid to achieving a balance of groups of precedents in datasets, analyzing possible artifacts when forming a sample (for example, the tendency to an increased probability of getting into the sample of individual cases — as in the comic survey about Internet access conducted on the Internet), monitoring that all significant factors fall into the training sample. Reasonable and systematic measures can reduce risks to an acceptable level."

Aleksandr Vecherin,
Associate Professor, Department of Psychology, Faculty of Social Sciences, Higher School of Economics

"AI developers are able to successfully filter offensive and downright negative statements. Unfortunately, many negative stereotypes do not contain the key features of such statements, which creates great difficulties in filtering such content. To solve this problem, it is required at the first stage to study existing biases from a linguistic and psychological point of view, identify characteristics associated with the most vivid emotional reactions of the user and develop a system of criteria for evaluating statements. These results can then be used to train models further."

08 The problem of accountability: using the example of medicine, what responsibility does an AI developer have in case of harm to a patient's health?

Answer:

The legal responsibility of AI developers is almost always regulated by industry legislation where the AI system is used — in this case, medical. If this issue is not resolved, the developer may be ethically responsible if known errors have been hidden, measures have not been taken to correct failures, or insufficient information has been provided about possible system risks. As a rule, the developer of the AIS does not bear ethical responsibility for the consequences of using the system if the risks and limitations have been explicitly and clearly communicated to medical professionals.

Recommendations

For developers:

1. **Use reliable sources of information** to create high-quality datasets for machine learning.
2. **Test and register the AI system** to confirm its safety and effectiveness based on the evidence collected.
3. **Ensure that AI systems are regularly updated to be aligned with new medical standards and research.** This will help minimize the risk of using outdated data and increase the relevance of the system in medical practice.
4. **Develop AI systems with the possibility of explanation if it does not conflict with the quality of the solution.** It is important that healthcare professionals can understand what data and logic the AI recommendations are based on, which will increase trust and reduce the risk of errors.

For medical professionals:

1. **It is recommended to maintain a constant dialogue between developers and doctors** about possible errors and limitations of the system in order to minimize risks for patients.
2. **Observe the principles of caution and reasonableness when making decisions using AIS.** Evaluate potential risks to understand when it is necessary to contact the developer for additional instructions.
3. **Integrate AI as an auxiliary tool, not a substitute for human analysis.** AI recommendations can be useful, but should always be considered as an addition to the clinical opinion and experience of the doctor.
4. **Check the data and recommendations proposed by the AI before using them in treatment.** It is especially important to do this in difficult cases or when there are doubts about the accuracy of the system's proposals in order to avoid incorrect decisions.

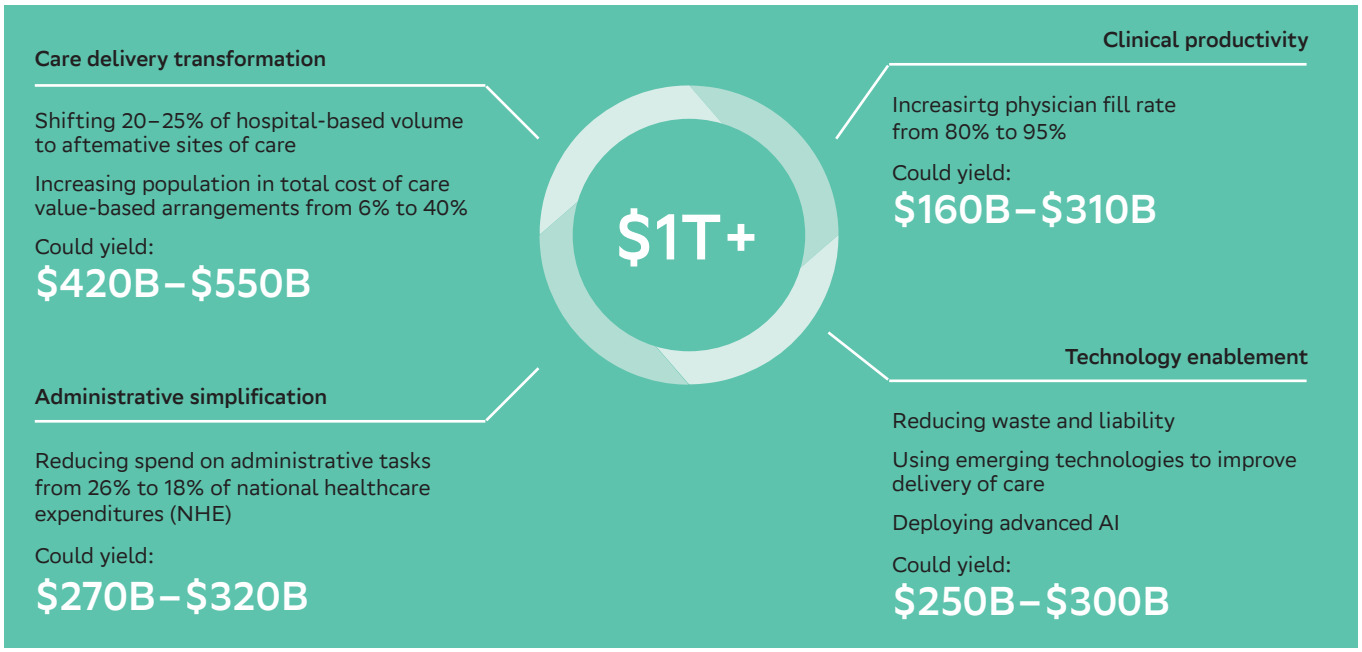
Justification

- According to a group of researchers from the UAE and Egypt, it is **unfair to hold developers responsible, since AI systems work autonomously, and not all errors can be foreseen or prevented at the development stage.** Manufacturers are responsible for defects that may be related to the design process or inappropriate instructions only if there is a foreseeable risk of harm associated with the product[77].
- Using AI to provide medical care **is not much different from using any other medical device.**
- **Poor-quality operation of these devices outside the declared characteristics may be associated with a malfunction** that the attending physician could not have foreseen or detected.
- According to a study by the American law firm Leeseberg Tuttle, **currently the responsibility for harm to the patient is still assigned to the medical professional who provided assistance**, regardless of the tools they used. Moreover, the study showed that the main cause is often human error[78].

Research on the issue:

According to a 2022 McKinsey study, **generative AI technologies represent a significant new tool that can help unlock some of the untapped potential of the medical industry**[79]. This is possible by automating tedious and error-prone operational work, bringing long-term clinical data to the attention of a doctor in a matter of seconds, and modernizing the infrastructure of healthcare systems. Joint investments in these areas can bring profits of $1 trillion to $1.5 trillion

Potential profit for the healthcare sector due to the introduction of AI



Source: McKinsey[79]

The European Union's approach:

In 2014, the Commission for the Ethics of Research in Information Sciences and Technologies (CERNA) proposed several recommendations on the ethical use of robots in medicine. They were published by the European Parliament[80]:

- Researchers should seek and follow the opinions published by the current medical ethics committees.
- Researchers working on robotic systems should strive to maintain the autonomy and control of the people for whom the technology is applied.
- Researchers must ensure that all actions by robotic systems remain reversible.

The approach of the Russian Federation:

A medical worker does not bear personal civil liability for medical care provided due to the provisions of Article 1068 of the Civil Code of the Russian Federation on the responsibility of a legal entity as an employer for the actions of employees.

According to Art. 1096 of the Civil Code of the Russian Federation, damage caused as a result of deficiencies in the service is subject to compensation by the person who provided the service. Thus, in a situation where medical AI equipment met all the requirements of certification and standards of medical care, harm requirements are imposed on a medical institution (as a person who uses certain equipment).

Research on the issue:

A study by French companies MACSF and WITHINGS shows that among 1,037 doctors that are MACSF members, 43% of those working with connected devices for diagnosis used them often, 30% always, 27% used them for remote tracking, and 25% used them for primary or secondary prevention[81].

In addition, **more than a third of doctors are still cautious about the applicable liability regime** if the treatment they recommended led to deterioration in a patient's health.

Practices:

1. In 2017, **a dental robot developed by Chinese specialists operated on a patient for the first time without the participation of doctors.** The robot successfully implanted two teeth previously printed on a 3D printer. According to the results of the operation, it was found that the implants were set with an error of 0.2–0.3 mm, which is acceptable according to medical standards. It was clarified that the decision to create a robot was made against the background of a shortage of qualified dentists in China[82].

2. Currently, neuroimaging tools require MRI scans with several requirements, including resolution and contrast for accurate 3D analysis. However, most MRI scanners around the world do not meet the required criteria. Therefore, researchers at Harvard Medical School have developed **the SynthSR AI system for converting low-resolution MRI images**. Such an improvement in image quality could revolutionize their use in critical conditions or in places with limited medical capabilities where there is no MRT1 equipment[83].

3. Researchers from the Breast Cancer Now unit at King's College London have created **an AI model to predict the likelihood of breast cancer spreading** in patients with triple negative breast cancer. The AI model, a deep learning platform called smuLymphNet, is used to analyze images of lymph nodes in cancer patients, comparing them with patient records and determining whether cancer has spread[84].

What do the experts think?

Anton Kiselyov,
Deputy Director on Science and Technology of National Medical Research Center for Therapy and Preventive Medicine, Moscow, Russia

"In the doctor AI link up, the role of AI currently usually consists of supporting medical decision-making, or at most it is used when a second opinion is needed. At the same time, the admission of AI to perform such functions in practical medicine is strictly regulated. AI services that are directly involved in the process of providing medical care are subject to mandatory state registration according to the rules applicable to medical devices. It is at this level that the delineation of the spheres of responsibility for the admission of AI services into clinical practice takes place. The responsibility for making a decision on a particular patient remains entirely with the doctor, regardless of whether they took into account the 'opinion' of the AI assistant or not."

Vyacheslav Shulenin,
General Director of the Moscow Center for Innovative Technologies in Healthcare

"Despite the limitless potential of using neural networks, it is impossible to completely replace specialists, because decision-making is a human responsibility. And no computer or system can be assessed as subjects of legal and ethical assessment of actions, as well as their consequences[85]."

Andrey Almazov,
Deputy Director for Project Activities of the 'National Medical Knowledge Base' association

"The question being asked requires clarification of what type of harm and in what way, due to the use of AI, it can be caused. For example, let's assume that the radiation dose of a CT has become regulated by AI, in which case the developer is undoubtedly responsible for safety – not just the AI, but the entire medical product is legally responsible. And, by the way, the presence of AI here does not change anything in comparison with current practice. The issue here is legal, rather than ethical.
Another polar hypothetical case could be where the AI interpreted a patient's lab results in such a way that it caused psychological trauma. This is really an ethical issue, but a similar situation can typically occur without AI – namely the deterioration in a person's physical or emotional state, unintentionally provoked by a medical professional. Who's in charge here? Apparently, the developer, because AI is not a subject, but becomes a co-participant in the process, invading relationships that previously remained only in the 'doctor–patient' circle."

09 The problem of delegating decision-making: in the case of the judiciary, will AI be able to replace a judge?

Answer:

In order to delegate AI decisions, it is always necessary to take into account the requirements of legislation and the positions of judicial authorities on this issue (if any). If the law allows such a possibility, from an ethical point of view, AI can delegate independent consideration of a small category of cases, where it is not necessary to take into account the subjective (psychological) side of the behavior of the participants in the process and assess their personality. In other cases, the role of AI may be reduced to the role of an assistant to the judge used for selecting and analyzing information.

Recommendations for the introduction of AI in the courts:

1. **First of all, evaluate in which cases the law allows the use of AI and how.**
2. **Ensure human control over the use of AI systems in the judiciary.** Solutions proposed by AI should be reviewed and approved by a judge or another responsible specialist in order to eliminate the automation of critical errors and maintain human control over the process.
3. **Only specialized closed AI models trained on verified data should be used.** The use of publicly available models trained on open data from the internet is unacceptable, as this can lead to erroneous conclusions and undermine confidence in the judicial process.
4. **The AI model and the data for its training must be verified by the professional community and the state.** This will help ensure that the model uses relevant and reliable legal positions that comply with legislation and judicial practice.
5. **When changes in judicial practice occur, the model should be updated in a timely manner**. Regular additional training on new legal norms and positions will ensure the relevance of the AI model and its adequate application.
6. **Program the models so that they can report a lack of data to make a decision.** The model should be able to inform users if there is not enough data for an informed conclusion, to avoid making incorrect decisions.

7. **At any stage of the application of AI in the judiciary, participants in the process should have the right to challenge the results of AI if it affects them personally.** The ability to review decisions made with the help of AI provides an additional level of protection for the rights of participants and makes it possible to eliminate mistakes.

8. **Addressing the issue of public confidence in the systems used requires transparency.** Disclosing information about the model used and providing access to its findings to participants helps to build trust and understanding of the work of AI in court cases.

9. **Most of the recommendations above are generally applicable to address the issue of the ethics of delegating AI decision-making.**

Justification

- **When resolving a case, the judge takes into account moral aspects** (for example, humanity, proportionality) and subjective factors of the situation (reasonableness, conscientiousness), which is in the emotional sphere of a person and which AI cannot cope with.
- **AI can make independent decisions on court cases in which the psychological aspects of the behavior of the parties are not studied**, and the decision is made without oral proceedings based on written evidence — as well as indisputable (court orders) and minor civil cases (small claims up to a certain amount).
- According to a study published by the International Journal of Judicial Administration, the algorithm's **solutions cannot be used independently as a prescription**. Thus, AI cannot be allowed to resolve issues of the defendant's guilt in criminal proceedings, since they are related to the assessment of the subjective side of the defendant's behavior[86].
- Researchers at Lexis Nexis, an international information services company, believe that **AI offers the shortest pathway to optimizing the process of analyzing court cases**. For other categories of cases, AI assistance with the selection and analysis of case information, the preparation of a case review forecast (predicative report) and the text of the judicial act can be very significant, but the decision is made by a human judge[87].

Regulatory approaches:

1. In December 2018, the first international act specifically dedicated to the use of AI in justice appeared — **the European Ethical Charter on the Use of AI in Judicial Systems**, approved by the The European Commission for the Efficiency of Justice. The Charter emphasizes the need to fully guarantee respect for human rights, the principle of equality of the parties, the presumption of innocence, transparency and non-discrimination in the use of AI technologies in the judicial system[88].

2. Guidelines on the ethical use of AI in judicial proceedings are also being adopted at a national level. For example, in July 2024, **a Guide on the use of generative AI for judges and judicial officials was developed in Hong Kong**[89]. As a result, judges and judicial service personnel can intelligently and responsibly use generative AI in the course of their work, where appropriate. However, it is prohibited to delegate judicial functions to AI — all court decisions must be made exclusively by judges.

3. In December 2023, **a Guide for all judicial office holders in courts and tribunals**[90] was also published in England. According to the Guide, AI can be useful for summarizing large amounts of information or performing administrative tasks, but AI is not recommended for conducting legal research due to the risk of "hallucinations" and factual errors.

Practices:

1. **In 2021, the Supreme Court of the People's Republic of China ordered judges to consult with AI when making decisions.** The Smart Court system, launched in 2015, automatically scans court cases for references, recommends laws and regulations, develops legal documents and corrects alleged human errors, if any, when delivering a verdict[91].

2. At the initiative of the French Ministry of Justice, **two appeal courts agreed in the spring of 2017 to test the Predictive justice** software for appeals. This AI tool, based on the analysis of civil cases of all French courts of appeal, offered a decision to judges, which was supposed to promote the principle of equality of citizens before the law[92].

3. **The Russian judicial system has also begun testing AI.** A pilot project has been launched in the Belgorod region: magistrates at three judicial districts have engaged AI to prepare court orders to collect taxes from citizens for property, transport and land. AI tools should help judges prepare documents, including creating a case file in the internal court system[93].

4. Judge Juan Manuel Padilla in Colombia considered a case on covering the costs for treatment and transportation for a child with an autism spectrum disorder. It was necessary to find out whether all expenses should be covered by insurance, since the child's parents could not afford them. The judge asked ChatGPT whether the child's family should be exempt from payment for treatment. The neural network replied that according to Colombian law, people with autistic disorder are exempted from paying for therapy. The court's decision was the same as the chatbot's response. At the same time, the judge said during an interview that he made the final decision independently and used precedents from previous rulings to do so. **Consulting with a neural network helped speed up the process**[94].

Having analyzed the research on the topic, the most popular ethical principles for AI in the judicial system are:

1. **The principle of respect for human rights**, by virtue of which the use of AI should not detract from the adversarial nature of the process and the right to a fair trial.
2. **The principle of quality and safety**, which involves the use of certified software evaluated by both technical specialists and lawyers.
3. **The principle of human control**, according to which the judge and the parties to a dispute should be able to disagree with the decision proposed by AI and challenge it.
4. **The principle of non-discrimination**, which includes a ban on the use of data that may lead to bias against certain groups of people.
5. **The principle of transparency**, by virtue of which all the features of the technologies used should be brought to the attention of citizens in an accessible form using understandable language.

What do the experts think?

Victor Momotov,
Chairman of the Council of Judges of Russia

"Artificial intelligence cannot become a guarantor of the protection of human rights and freedoms and ensure fair and humane justice. Therefore, its application is possible only in a limited form, with clearly defined limits and rules. The interaction of judges and court staff with artificial intelligence technology should lead to synergy, while maintaining the dominant human role in the partnership. As information technologies develop, their scope of application expands from technical and routine functions to solving more complex tasks, and information systems become an environment for the implementation of procedural actions[95]."

Elena Avakian,
Vice President of the Federal Chamber of Lawyers of the Russian Federation

"The use of AI in justice depends on the area in which the dispute is being considered. A judge may categorically not be replaced in criminal proceedings. Because we judge a person, their act, their subjective side. So, allowing a machine to make judgments about human action means losing the species competition. As for the administrative-legal and civil-legal spectrum, there are already cases of writ and simplified proceedings, where AI not only can, but must replace judges. Here, AI will work to strictly apply regulations in specific conditions."

Anatoly Vyborny,
Deputy Chairman of the Committee on Security and Corruption Control

"Artificial intelligence can be painlessly entrusted with all tax disputes, as well as challenging government decisions — for example, decisions by road traffic control investigators. I emphasize that we are talking about small amounts and simple disputes — that is, in this case we are talking about tax disputes and administrative penalties for violations of traffic rules[96]."

Andrey Neznamov,
COO of the Human-Centered AI Center, Sberbank, Chairman of the Ethics Commission in the field of AI

"It is important that AI helps to free up courts by performing automatic, routine tasks, and those in which the use of AI would reduce the number of errors. However, all this must strictly comply with the procedural rules of a particular country. Therefore, it is important for us that regulators gradually create a procedural framework for the implementation of AI[97]."

10 The social rating problem: is it ethical to use AI to create a social ratings?

Answer:

The main ethical issue is not the application of AI, but rather the application of the social rating itself — and this issue is extremely debatable. Nevertheless, research shows that the ethical prerequisites for the application of the social rating system are preliminary discussion with the public and transparency of the application of the rating system.

Research recommendations:

1. **Before implementing a social rating system, conduct a multi-stage public discussion** with the participation of experts, human rights defenders and representatives of various social groups. Spreading a social rating that affects all spheres of life without universal discussion and agreement can be considered unethical.
2. **It is important to develop an ethical and legal framework** for regulating the social rating system. On this basis, create institutions of public control — so the potential of new technologies can be realized without compromising fundamental human rights.
3. **It is important to ensure the transparency of the social rating system.** Users need to provide information in an understandable form about which of their data can be used and which data can be considered publicly available, as well as the impact of this information on the rating.
4. **The possibility of appeal is important.** Everyone should be able to find out their rating, as well as challenge its correctness or consequences, if necessary, in order to prevent unfair sanctions or mistakes.
5. **It is necessary to create a data protection system that will guarantee the confidentiality and security of personal information.** The data used must be protected from unauthorized access to prevent abuse and leaks.
6. **It is advisable to regularly audit and evaluate the social rating system.** It is important to ensure that the system remains fair and does not infringe on the rights of individual groups of the population. Regular independent checks will help identify possible risks and shortcomings in the rating.

Justification

- Scientists at the Middle East Technical University of Turkey claim that **the social rating system pursues legitimate and socially-useful goals**. Data should be collected and analyzed from various sources in order to create a secure and so-called trust-based society[98].
- Researchers at Vladimir State University emphasize that **the use of social rating without proper legal regulation can violate the privacy of citizens**. From the point of view of protecting the right to privacy, it is important that various kinds of personal information are used as sources for social rating[99].
- A group of researchers from Israel and Japan believe that **the greatest risks are associated with opacity of the social rating system**. It is not always clear which factors and how much they influence a person's score. As a result, certain forms of control can have a real impact on people's lives due to the 'prejudices' that have developed in the system, as well as the errors within it[100].
- **Social rating carries risks of collision of different societies and a loss of personal freedom.** The social environment is represented by a variety of societies: traditional, conservative, religious, technocratic, avant-garde, often with opposing values and beliefs, which requires an approach of 'ethics of discourse'.

Social rating is a system for monitoring the social activities of citizens, which is evaluated according to several parameters. Based on the assessment, which is a rating score calculated using special algorithms for digital processing of a set of certain data, a range of opportunities and services that a particular citizen can use is formed. Depending on the number of criteria approved by the system, the rating score is set: the higher the score, the more privileges and opportunities a citizen has, and fewer restrictions[101].

The EU's position:

The European Union Regulation on Artificial Intelligence prohibits the use of social rating systems[102]. According to EU legislators, AI systems that provide social assessment of individuals by public or private entities may violate the right to dignity and non-discrimination, as well as the values of equality and justice. The social assessment obtained with the help of such AI systems can lead to negative consequences that are disproportionate to the severity of human behavior.

UNESCO's position:

UNESCO's Recommendations on the ethical aspects of AI enshrine the principle of human final decision-making in cases where it is assumed that decisions have irreversible consequences or those that are difficult to reverse, or when the decisions may relate to issues of life and death. In particular, artificial intelligence systems should not be used for social assessment or mass surveillance[103].

The social rating system in China:

- **The most extensive example of social rating use is implemented in China**[104]. It covers more than 1 billion people and takes into account 160 thousand different parameters, including factors such as credit history, compliance with laws, timely payment of bills, volunteer activity and even statements on social networks.
- Authorities or companies rate a person's social behavior from 0 to 1000 or from A to D. **The rating is based on information** received from official sources — tax office, police, government agencies, and also from digital sources: search history, online purchases and social media activity.
- **Citizens with a low social rating are included in a 'black list'.** Such citizens may be denied a loan, mortgage or admission of children to a private school. It is important to note that it is possible to be removed the blacklist, for example if a person engages in socially important activities.
- **Proponents** of social rating claim that it can make society safer, fairer and more efficient by encouraging people to behave more responsibly and ethically. **Critics** see it as an instrument of absolute control, violation of privacy and restriction of freedom.

A similar experiment in Venezuela:

Venezuela's smart card, known as **the 'national card', collects a variety of information about cardholders and stores it in a government database**, which the government claims will help them provide citizens with better services. The database, according to employees of the card system, stores a variety of information, including medical history, social media presence, political membership and whether a person voted[105].

Research on the issue:

1. In the period from February to April 2018, German researchers in collaboration with Chinese companies conducted a nationwide online survey to identify **how the behavior of Chinese citizens changed after the introduction of the social rating** system[106]. More than 350,000 Chinese people participated in the survey.
   The results show that the majority of respondents (94%) reported a change in their behavior. The reported behavioral changes were often caused by a desire to improve personal performance: 91% of respondents changed their behavior at least once to positively influence their rating (for example, they participated in charity). Meanwhile, 85% reported that they changed their behavior at least once in order to avoid penalties/restrictions (for example, they carefully followed the rules of the road).

2. In January 2024, German scientists conducted **a study on the degree of acceptance of social rating systems by citizens of countries from Southeast Asia**, based on the Chinese system itself. Among the respondents, 50% would fully or to some extent approve of the introduction of a social rating system in their countries, while only 15% were strongly or to some extent against the introduction of such a system[107].

What do the experts think?

Xue Lan,
Cheung Kong Chair Distinguished Professor, Dean of Schwarzman College and Dean of Institute for AI International Governance of Tsinghua University

"The system is undergoing significant changes, it is still at the testing stage. It should be borne in mind that China's population is 1.4 billion people, and there are many problems that need to be solved. Reports that 'big brother' is trying to take everything away from everyone are not true... I do not see that the social rating gives China any special advantages, I do not think that the government uses it to obtain any commercial benefit. There is no evidence of this in China[108]."

Roman Dushkin,
the general director of AI-developer 'A-Z expert'

"If such a system is implemented and if it is extended to cover everyone, this will instantly lead to the stratification of society. Society is already divided into layers and strata, and the use of a social rating system will only exacerbate this further[110]."

Holger Zscheyge,
founder and managing director of Infotropic Media, co-founder of Moscow Legal Hackers, Ambassador of the European Legal Technology Association (ELTA)

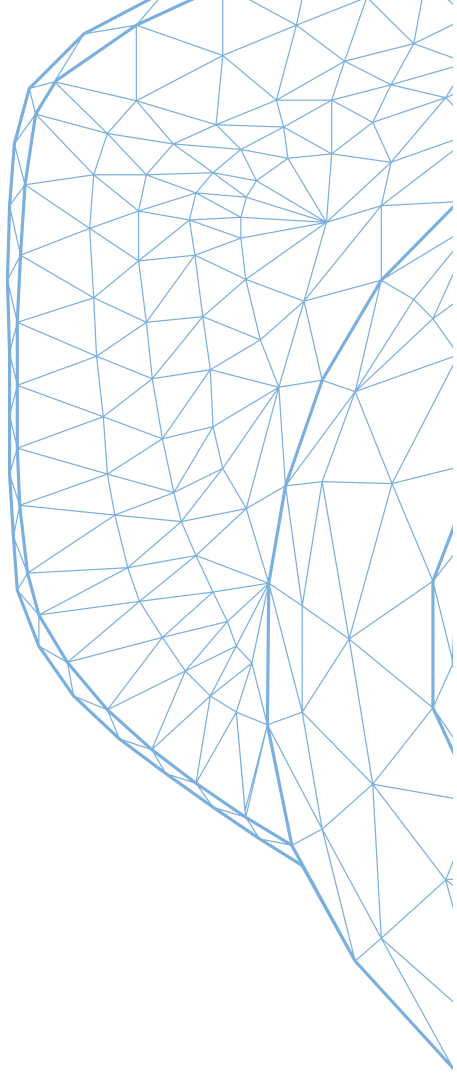
"Social scoring in China is a separate and unique case. This is not an attempt to minimize commercial risks, it is a method of total control over the population. This is what the authors of books and feature films have warned us about, ever since Orwell. The problem with social scoring is that the state can arbitrarily set parameters and thus make life a living hell[109]."

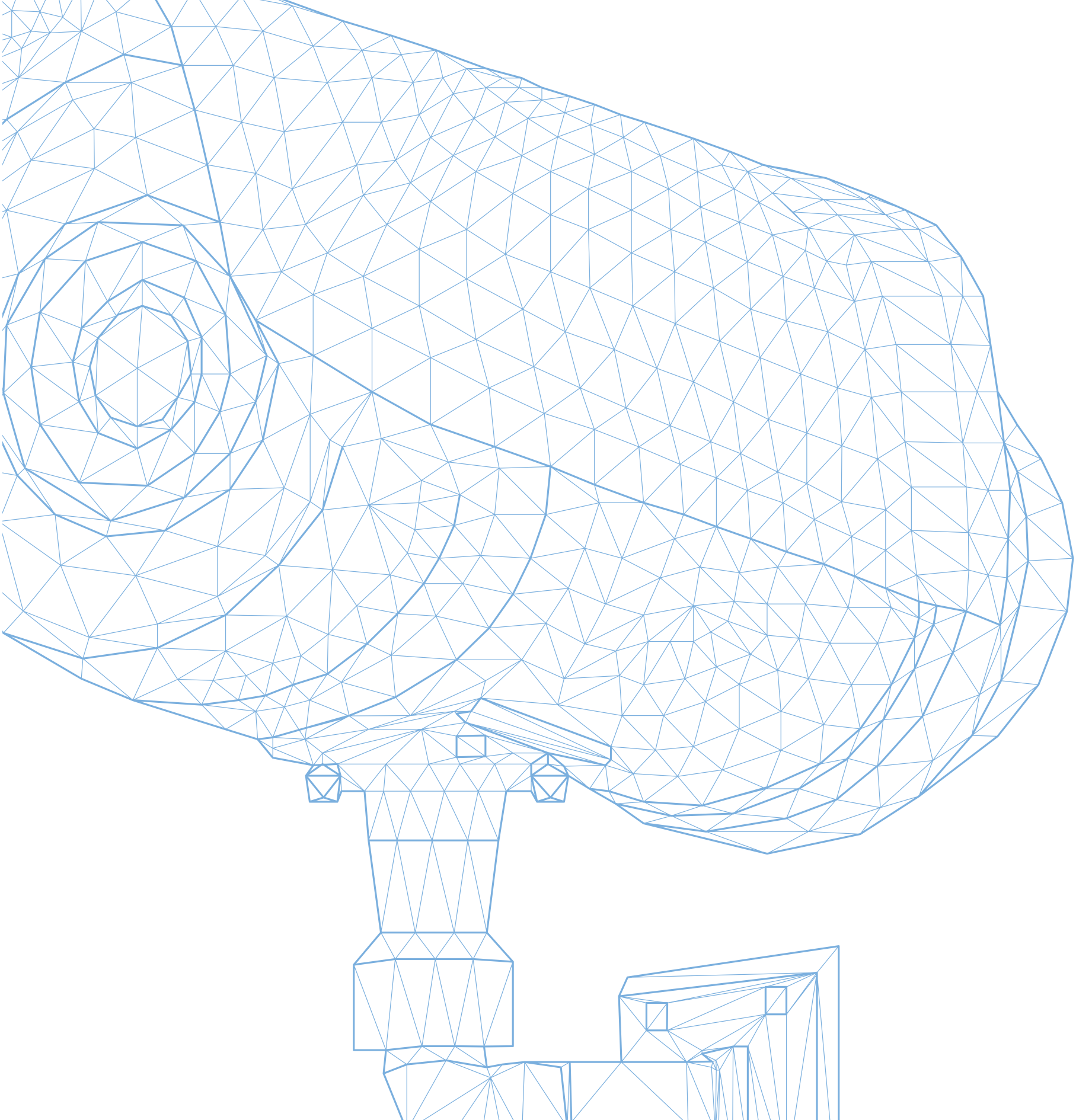
Andrey Svintsov,
Deputy Chairman of the State Duma Committee on Information Policy, Information Technology and Communications

"The introduction of scoring will lead to massive burnout of people, especially young people, who will strive to have a high rating. In the end, we will get a nation not of people, but of robots. In modern Russia, even with the highest level of artificial intelligence development, such systems are inapplicable and, I think, unacceptable[111]."

Chapter 02

AI and Confidentiality





11 Is it ethical to use personal data for AI learning?

Answer:

It is ethical only taking into account compliance with the requirements of the legislation on the confidentiality of personal data and respect for the rights of data owners and only in cases where this is necessary.

Justification:

- The OECD, in its report 'AI, Data Management and Privacy'[112] recalls **that just because data is available does not mean that it can be collected and used to train AI models**. Personal data (PD) must be obtained legally and in a manner in which its use is compatible with the original purposes.
- **Personal data (PD) can be used for scientific research, including in the interests of society as a whole.** For example, in medicine: to study new treatment methods and develop medications.
- **You cannot use data collected for another purpose to train AI.** The use of personal data in machine learning is an independent application for data processing and requires a legislative basis[113].
- The office of the UK Information Commissioner warns that machine learning **models trained on personal data may inadvertently increase discrimination**. For example, data from the resumes of past applicants for training the AI system used in hiring may contribute to gender discrimination, since men have long been considered more suitable candidates for certain positions[114].
- According to a study published by heyData, **PD in machine learning is used to improve the quality and efficiency of digital services.** The use of advanced PD protection methods allows you to balance the confidentiality and the accuracy (usefulness) of a machine learning model[115].

Recommendations for developers:

1. **Take a responsible approach to decision-making about training an AI model based on personal data.** If there is no clear understanding on how PD could help training a model and what results it can lead to, then it is better not to use PD.

2. **Receive the user's consent to use PD in machine learning, if required to do so by law.** If consent is not required, it is considered ethical to warn the user in any form about the use of PD in machine learning.
3. **Exclude the possibility of unauthorized access to PD as a result of training the model.** Use anonymization and data encryption methods; limit the number of people who have access to PD; conduct regular monitoring and audits on the system to identify potential threats and security breaches.
4. **Identify sensitive data (religious beliefs, sexual orientation, mental illness, etc.) that could lead to unjust outcomes, and their significance (weight).** Assess the fairness of a machine learning model both from the point of view of the interests of an individual and social groups.

Research on the issue:

According to an IBM study, **federated learning** is a learning method that allows you to configure a centralized machine learning model without data exchange which significantly raising the level of confidentiality[116].

In the federated learning system, each device has its own copy of the model. These devices train their own copy of the model using their data. Then they send the parameters of their models to the main device or server. There, these parameters are combined and the overall model is updated. This process is repeated until the desired accuracy is achieved.

Thus, the idea of federated learning is that training data is not transmitted between devices or between parties. Only updates related to the model are transmitted.

Practices:

1. Clearview AI has collected billions of images from social media without users' consent and created a facial recognition system to sell to law enforcement agencies and private companies. **Since the photos were obtained without permission, many countries have recognized this practice as illegal.** Clearview AI has faced numerous lawsuits, as well as a ban on its activities in some countries (for example, in Australia and France)[117].

2. In the spring of 2023, the Italian data protection authority restricted access to ChatGPT **due to a leak of user data**[118]. In addition, OpenAI had not notified users that it was collecting their data to train algorithms. As a result, the GDPR requirements on the legal basis for the processing and storage of personal data was violated.

3. In May 2024, **Meta announced that it would use users' personal data to train AI**, particularly the photos published and made publicly available on the company's services. Users were given the opportunity to opt out of using their personal data for training. However, the mechanism for opting out is quite complicated. Users must fill out a long form specifying a detailed reason for their refusal.
On June 6, 2024, 11 complaints against Meta were filed in courts across Europe. In response, Meta publicly accused the plaintiffs of obstructing the development of generative AI [119].

What do the experts think?

Artyom Sheikin,
First Deputy Chairman of the Federation Council Committee on Constitutional Legislation and State Building

"The ethics of using personal data for AI training largely depends on compliance with a number of principles: legality, consent, confidentiality, compliance with the purposes of use, as well as the willingness of AI developers to be responsible for their actions. Thus, this process must fully comply with the current legislation in personal data, citizens must consent to the use of their data, as well as be informed about how it will be used and for what purposes. In addition, the data must be protected from leaks and depersonalized in order to reduce the risk of its use for illegal purposes."

Eduard Lysenko,
Moscow Government Minister, Head of the Department of Information Technologies

"Depersonalized personal data is extremely important for AI training. For example, a medical decision support system will not be able to tell a radiologist that a tumor is suspected in a specific CT scan in a specific area of the lung unless this system has already been trained using thousands of images. At the same time, in order to train the system, it does not need to know who each of these images belongs to — it only needs to learn how to recognize malignant tumors. The situation is similar for systems in other areas — education, transport, ecology, etc. [120]"

12 Is it ethical to collect user data from a smartphone or smart device for AI training?

Answer:

It is unethical if data is collected without complying with the law, in particular, informing the user and obtaining consent to the collection and processing of data; in the absence of such legislation — it is considered unethical if done without warning the user.

Justification:

- **The collection and processing of user data is regulated by legislation on personal data and communications.** Usually, users must be informed about the processing of personal data and provide consent.
- UNESCO warns that **data collected by the IoT (Internet of Things) is easy to combine to create a highly accurate human profile**. Even if personal data was collected in accordance with legal requirements, the volume and diversity of such data may lead to a threat to confidentiality[121].
- The Office of the Information Commissioner of the United Kingdom believes that **only necessary data can be processed by default**[122]. That is, data that are necessary for the normal functioning of the device.
- Researchers at the All-Russian State University of Justice state that **projected 'customer-focused orientation'** reduces the risks of interference in the user's privacy by ensuring data confidentiality and transparency in their collection[123].
- **Leakage of user data increases the risk of negative consequences for the user.** In this case, personal data can be used for blackmail, fraud, etc.
- **Data obtained as a result of depersonalization is personal.** Such data potentially makes it possible to identify a person if additional information is available or using certain analysis methods.

Recommendations for developers:

1. Study the legislation governing the protection of personal data, personal life and privacy of correspondence. This will help to analyze and (if possible) prevent the legal risks.

2. **Obtain the user's consent to the collection and further processing of user data**, if necessary, in accordance with legal requirements. Such consent must be informed.
3. **Integrate privacy protection tools into the product functionality.** For example, notify the user about the collection of information and provide the opportunity to restrict or prohibit the collection of data 'in a manual mode'.
4. **Minimize the collection of user-identifying information** if it is not required for the normal use and operation of the service.
5. **Provide the user with the opportunity to receive information about the data used.** It is important for the user to know what data you are collecting and how you are using it.
6. **Use various methods of anonymization and encryption of personal data if transferring it to third parties.** This will help prevent any consequences of data leakage.

Recommendations for users:

1. **Review the company's privacy policy.** It should specify what data is collected, how it is used, and how the user can control their data.
2. **Use the product's functionality to protect your data.** If the device obviously does not require a microphone or camera, geolocation, etc. for its normal operation, limit data collection permissions in the settings.
3. **Send error information to the developer to fix bugs.** This will help optimize the operation of the device/application and increase user satisfaction with the product.
4. If technical support could not assist in resolving your issue, contact a higher authority. For example, special commissions that specialize in corporate ethics, judicial authorities or other government agencies

Research on the issue:

1. According to a survey by the Russian Public Opinion Research Center, more than a quarter of Russians (28%) use 'smart' devices for home at one time or another [124].
   Russians consider the main threat of this to be the possibility collected data being transferred to third parties (15%). Another 6% replied that the collection and analysis of user information is an invasion of privacy, violation of rights and freedoms, 5% do not exclude the possibility of surveillance or espionage using 'smart' devices. 23% of smart device users take actions that prevent the collection of personal data. The most frequent measures taken to protect privacy are covering a laptop's webcam (12%) and disconnecting devices from the network (6%).

2. Voice-activated assistants analyze every sound in order to recognize the activation phrase. **Similar sounds can often cause a false activation.**
Researchers from Unacceptable have discovered more than a thousand phrases that lead to the activation of the Alexa, Google Home, Siri and Microsoft Cortana voice assistants[125]. For example, Siri responds to the word 'city', while Cortana answers to 'Montana'. Many of these phrases are found in movies and TV shows like 'Game of Thrones', 'House of Cards' and on the news.
Moreover, Siri can be activated by the sound of a zipper or by raising your hand.

Practices:

In 2024, news broke that **Google had suffered a huge data leak affecting** service users[126]. The Google Audio function had performed unintentional recording of children's voices, Google Street View had decrypted and saved car license plates, and the Google-owned Waze service had disclosed the home addresses of users.

What do the experts think?

Elena Suragina,
head of the working group on best practices for emerging ethical issues in the AI life cycle, AI Commission for Implementation of the Code of Ethics in AI

"In a vacuum, collecting user data with smart devices is not unethical if such collection takes place with the consent of the user. However, openness in user interaction and transparency of information are important in this case. Users should be aware that data is being collected and how this data will be used. Such openness should become a foundation for users' trust in digital services and in the business as a whole."

Andrey Kalinin,
CEO of MTS AI

"There are quite a lot of aspects to this issue. First of all, the ethics of collecting user data from various devices is a matter of compliance with applicable legislation and generally accepted standards. If there is an agreement from the user and they are aware of what data will be collected and where it will be used, then this is completely ethical. But before that, you should definitely make sure that the purpose for which the data is collected is consistent with the company's values and ethical principles. In addition, it is necessary to ensure the protection of data from potential leaks[127]."

13 Is it ethical to use AI in mass video surveillance?

Answer:

It is ethical to use AI in mass video surveillance systems to ensure public safety while respecting human rights, using it in cases and in accordance with the procedure established by law; in the absence of legislative regulation, such use of AI would be ethical with prior warning to citizens.

Justification:

- Korean scientists at Gachon University note that **mass video surveillance serves as an early warning and notification system in case of a threat**. AI from recording devices creates dynamic means of public safety — information is analyzed in real time[128].
- The OECD in its report 'AI and society'[129] reminds that **law-abiding citizens are not in danger**. AI video information processing focuses the attention of law enforcement agencies on security threats and offenses.
- Researchers at the University of Manchester claim that video **surveillance systems do not violate the right to privacy**[130]. Because they are located in a public place and are in the public interest, that is, to ensure safety and protect public order.
- **The use of AI in video surveillance increases the sense of security in public places.** It will make it possible to quickly analyze a situation, determine the algorithm of solutions and call the necessary service personnel, to ensure that human rights are respected.
- Scientists at the Manav-Rahn International Research Institute believe that **AI conserves human resources and optimizes the work of law enforcement agencies**. The use of AI will solve the problem of ensuring security in public places without involving a large number of service personnel, since it will make it possible to identify threats remotely[131].
- Judicial practice confirms that **mass surveillance cameras record video of passers-by in streaming mode from a distance.** As a general rule, this does not involve the processing of biometric data, unless identities are established[132].

Research recommendations

1. **Promote transparency and openness.** It is recommended to provide citizens with information about video surveillance systems: the goals, mechanisms and advantages of their use. This will help to strengthen public trust and ensure control over possible abuses.
2. **Regularly evaluate the effectiveness and impact of video surveillance systems on public safety and citizens' rights.** If necessary, implement improvements and adjustments to the system.

Recommendations for citizens:

1. **Check out the legal acts governing this issue**. Knowledge of the legal framework and your rights will allow you to better understand and objectively assess the effectiveness of mass video surveillance.
2. **Maintain your awareness of new technologies and their applications.** The authorities can hold public discussions and consultations on the issue of urban video surveillance — take part in such events.

Practices:

The use of artificial intelligence significantly increases both the speed of solving crimes and the percentage of resolved cases.

1. The **US** police uses AI to compare a photo of a person who has committed an offense with photos already available in the police database[133].

2. In the **UK**, facial recognition technology is used in real time. When a person passes under a surveillance camera, their image is automatically matched with images of wanted criminals[134].

3. The **French** Constitutional Council, having agreed to the limited use of AI at the Olympics, stated that the new measures could only be applied at sports, entertainment or cultural events in order to "prevent public order violations." The law will remain in effect until March 2025. France is the first country in the EU to allow the use of AI for surveillance[135].

Positions of international organizations and regulators:

- **UNESCO's Recommendations on the ethical aspects of artificial intelligence**[136] reinforce the following approach: in cases where it is assumed that decisions taken have irreversible consequences or are difficult to reverse, or may be related to life and death decisions, the final decision must be made by a person. In particular, artificial intelligence systems should not be used for social assessment or mass surveillance.
- **The European Union's AI Regulation** follows the same approach[137]. Prohibited applications for AI include making available for resale, commissioning for a specific purpose, or using artificial intelligence systems to create or expand facial recognition databases through inappropriate extraction of facial images from the internet or video recordings from surveillance cameras.

What do the experts think?

Sergey Sobyanin,
Mayor of Moscow

"There were a lot of skeptical comments about video surveillance in the city, all sorts of insinuations that it was bad that someone could be followed. The video surveillance system primarily, of course, works for the safety of the city. To date, 7,713 people on the federal wanted list have been detained in the metro and in the city[138]."

Vladimir Tabak,
General Director of ANO 'Dialog Regions'

"The use of artificial intelligence in video surveillance systems is a great opportunity to significantly improve the work on public order protection and crime prevention. But the issues of maintaining confidentiality, proper handling of personal data and the risks of misuse of such technologies are very relevant. It is important to ensure that video surveillance systems equipped with artificial intelligence do not invade people's personal space. Citizens should be aware of the video surveillance and, if possible, give their consent. It is worth considering the transparency of the goals of using AI, and data processing and liability procedures in cases where errors occur in the operation of systems."

14 Is it ethical to use AI to predict and prevent crimes?

Answer:

It is ethical to use AI to predict and prevent crimes, first of all, only in cases provided for by law, and while respecting human rights and freedoms.

Justification:

- Interpol and UNICRI in their joint report 'Towards responsible AI innovation'[139] note that **AI offers law enforcement agencies huge opportunities to prevent crimes.** Predictive policing allows you to identify the areas with the most criminal activities, plan patrol routes and efficiently allocate resources.
- **The multiple increase in financial transactions in the digital environment, as well as the transfer of confidential information, creates new threats.** The use of AI reduces the likelihood of errors related to the human factor, such as inattention or lack of qualifications.
- **AI is often used to counter cybercrime.** Today, machine learning methods are used to monitor the activities of an information system and a person in order to identify potential deviations, predict malicious applications and sites.
- A study by the US National Institute of Justice shows that **AI makes the work of law enforcement agencies more efficient and less dependent on human factors**. The use of AI in the processing of personal information makes it possible to increase the speed of processing, as well as reduce the risks caused by human inattention[140].
- **Predicting crimes using AI is based not only on profiling of people.** For example, according to the AI Act, customs authorities in the EU are allowed to use AI to predict the probability of detecting drugs or counterfeit goods based on known trafficking routes[141].
- Scientists from Marian College in India warn that **the use of AI in predictive policing poses a threat of discrimination**. The data may be erroneous, incomplete and biased due to the fact that historically or due to regional characteristics, individual social groups may be more often represented as criminals[142].

Recommendations for developers:

1. **Ensure that fundamental human rights are respected.** These include: the right to confidentiality, access to information, non-discrimination and appeal against unfair decisions.
2. **Provide a plan to mitigate possible risks.** The development and training of AI for use in forecasting and preventing crimes should exclude the possibility of discrimination on any basis, as well as the creation of false information.
3. **It is recommended to install a human control system.** Errors are possible when using AI, therefore, it is necessary to verify the decisions made by AI and take into account all the circumstances and evidence.

Research on the issue:

Indeed, initially, the use of AI technologies to predict crimes was quite controversial, because these systems did not take into account the prejudices that had developed over a long period in the work of law enforcement agencies (for example, cases of false prediction of recidivism by black populations in the United States are widely known).

Nevertheless, according to recent research by sociologists from the University of Chicago, the latest AI systems for crime prevention **can predict future offenses a week in advance with an accuracy of about 90%**[143]. The new model isolates crime by considering the temporal and spatial coordinates of discrete events and identifying patterns to predict future events. It divides the city into several regions and predicts crime only within a given territory, and does not rely on traditional district boundaries or political boundaries, which are also subject to change.

Practices:

The Ministry of Internal Affairs of the Russian Federation plans to introduce artificial intelligence into law enforcement activities.

In 2024, the agency is conducting research and preparing datasets for training and testing neural network models, and plans to develop two AI-based systems in 2025: ‘Clone’ and ‘Conjuncture’. The ‘clone’ will make it possible to identify cases of video image forgery while ‘Conjuncture’ should predict negative events and emergencies and simulate scenarios for responding to them.

These initiatives are included in the plan for the introduction of AI technologies into the activities of the internal affairs office of the Russian Federation for 2023–2025. The plan was approved by Deputy Minister of Internal Affairs Vitaly Shulika[144].

What do the experts think?

Alexey Minbaleev,
Head of the Department of Information Law and Digital Technologies at the Kutafin Moscow State Law University

"No matter what difficulties arise when using AI in countering crime, the state is unlikely to abandon its use in this direction. Any opportunity to restore the rights and legitimate interests of a person that have been violated as a result of crime must be realized. But at the same time, it is important to ensure control over the decisions made by AI."

Temirlan Salikhov,
specialist in digital forensics

"The rational use of artificial intelligence-based tools has provided new opportunities and significantly optimized the activities of specialists in digital forensics. The ability to process billions of data makes it possible to make critical decisions in a short time and dramatically affects public safety. It is important to remember that the final decision rests with a competent professional."

15 Is it ethical to use AI for scoring in retail, finance and other specific applications?

Answer:

Yes, the use of AI scoring in finance and retail is ethical, while respecting the key principles: non-discrimination, transparency, personal data protection and expert control.

Justification:

- Researchers at YABX Technologies (a financial institution in The Hague), say that AI expands the possibilities of personalization and increases the efficiency of service provision. For example, machine learning algorithms can identify patterns and trends that people or traditional assessment models may overlook. Such an adaptive approach not only improves the accuracy of assessment, but also allows for real-time adjustments, ensuring that the assessment system remains dynamic and responds to changing external conditions[145].
- A study conducted by scientists from several US universities highlights the importance of ensuring transparency in the application of scoring in banking and retail. Consumers should have the right to know how these systems work, understand what types of information are used, and how the algorithms of the AI model work[146].
- Scientists at the American National University state that ensuring fairness and avoiding bias is one of the most important tasks for credit scoring. AI models can be designed to minimize discriminatory factors and promote equity by focusing on appropriate financial behavior rather than demographic characteristics[147].

Recommendations for business:

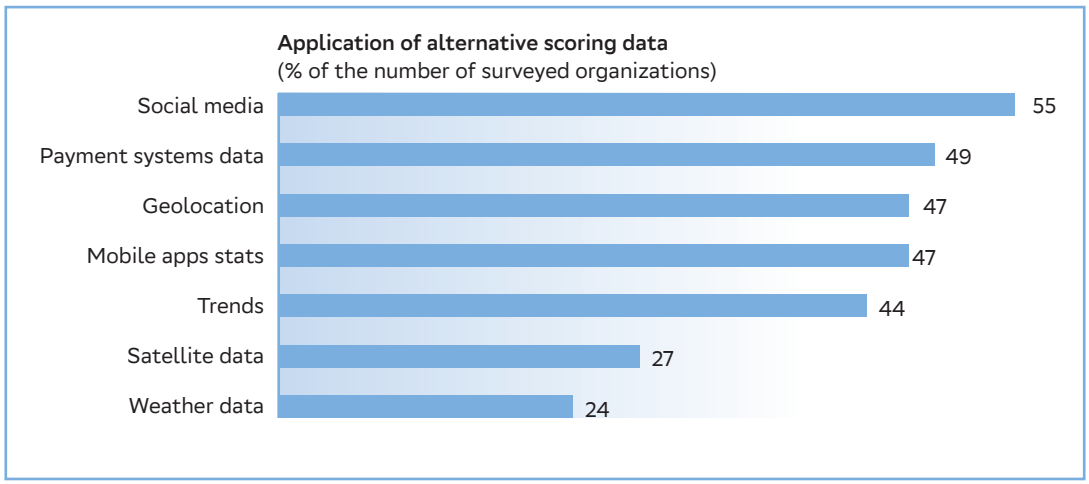
1. **Industry standards should be developed to ensure the ethical use of AI scoring in business.** These standards should enshrine the principles of non-discrimination, transparency and data protection.
2. **It is recommended to create mechanisms to explain the logic of the decisions made.** This will increase the transparency of AI scoring systems and the general awareness of users about the principles of their work.
3. **It is necessary to ensure the ability of the AI system to comprehensively consider the individual characteristics of customers.** The use of scoring in banking and retail should not lead to restrictions on the access of certain groups of the population to basic financial and consumer services.

4. **Support the creation of ethics commissions in the company.** These commissions will be able to evaluate the work of AI systems from the point of view of compliance with the principles of ethics, as well as to investigate cases of violations and take appropriate measures to protect the interests of users and customers.

Research on the issue:

In 2023, the Bank of Russia published a report on the Application of Artificial Intelligence in the Financial Market[148].

The report says that the use of AI by banks can be considered as an opportunity to further improve the efficiency and quality of their services, including by reducing costs, speeding up processes, resource optimization and processing large amounts of data.



Source: The Bank of Russia[148]

With the help of 'smart' scoring, creditors can analyze not only financial information about the borrower, but also 'alternative' data, according to the Central Bank report. The regulator attributed these indicators to:

- information from social networks;
- payment system data;
- geolocation;
- mobile application statistics.

What do the experts think?

Andrey Cherkashin,
Chairman of the Far Eastern Sberbank

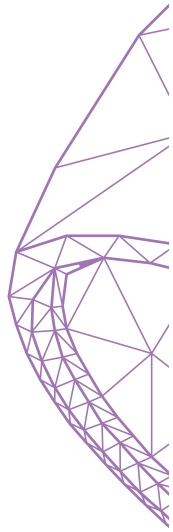
“Artificial intelligence finds the most active use in credit scoring[149], that is, in assessing the solvency of a person who wants to get a loan. In this case, the AI processes a large amount of data in a matter of seconds, analyzes thousands of parameters and makes a decision.Of course, the use of AI is not limited to issuing loans. In many of our business processes, we use models created with the help of artificial intelligence: from real estate search and transactions, to the introduction of AI into the work of a chatbot that analyzes speech, classifies calls to virtual assistants, verifies scans of documents[150].”

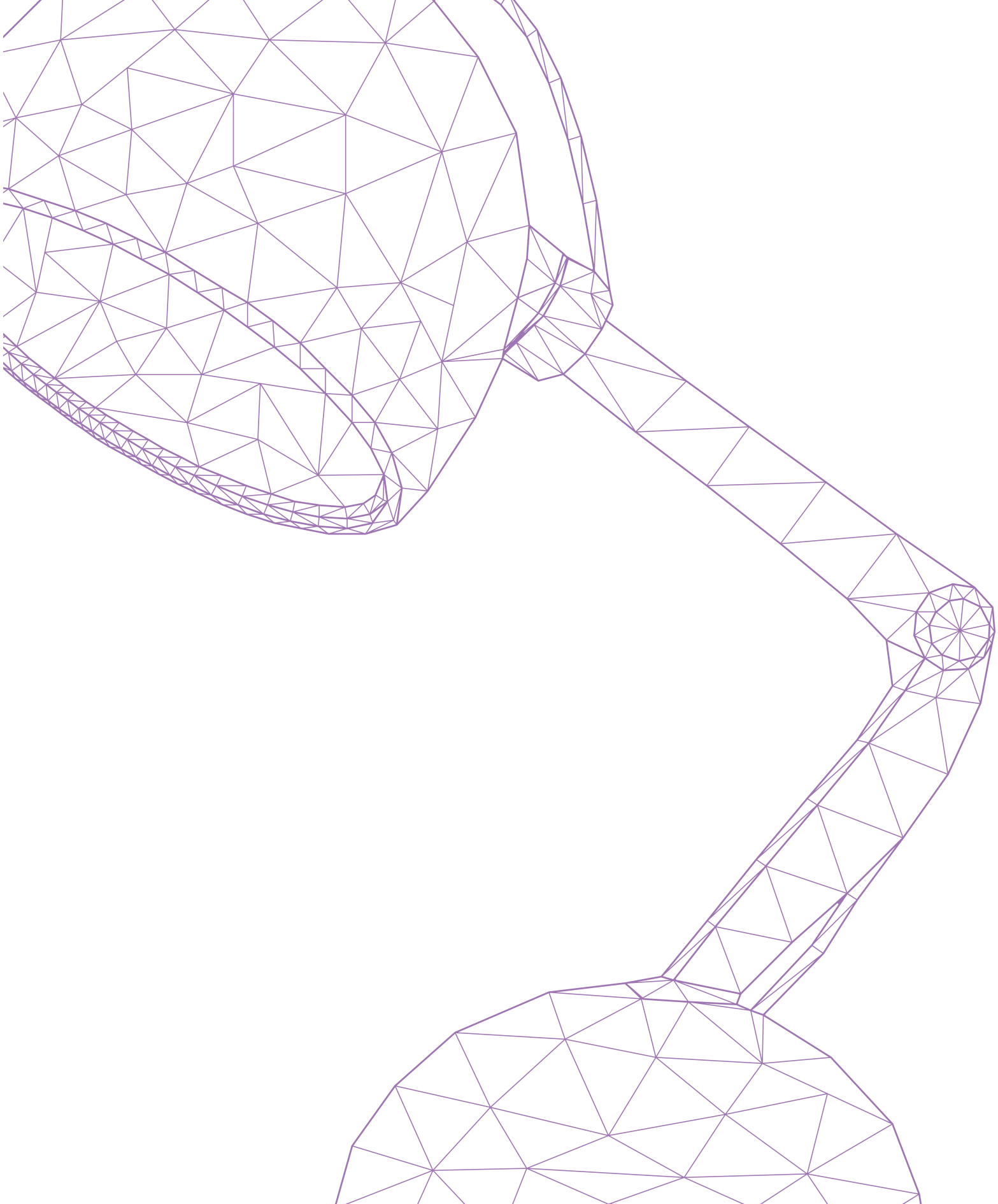
Anna Kazakova,
Risk Director, Vice President of T-Bank

“The ethics of using machine-leaning (ML) scoring in finance and retail depends on the context and goals. If machine learning helps to improve customer service and ensure fair lending, then it can be considered ethical. At T-Bank, for example, when calculating the credit limit, we use statistics to assess the solvency of customers, which helps protect them from financial illiteracy. At the same time, there must be measures to protect data and prevent abuse, so that privacy and user rights are a priority. It is important that algorithms do not reinforce bias and do not discriminate against certain groups of people[151].”

Chapter 03

AI and integrity





16 The learning challenge: how to avoid AI learning based on false information?

Answer:

Developers, data providers and customers that are implementing AI struggle to overcome the problem of false information in various ways: including by verifying data for compliance with legislation, testing the model and then retraining it.

Justification:

- In accordance with the European White Paper on AI, **various actors are involved in the development of the AI model — each struggling in its own way with unreliable information throughout the chain**. The most important are datasets, as they strongly affect the quality of the model [152].
- **Using a variety of data can help improve model accuracy.** Different data is used at the pre-training stage, since only high-quality data will be used at the fine-tune stage (final setup and training), taking into account the requirements of legislation (on personal data, trade secrets, intellectual property, etc.).
- According to a group of Russian and Austrian scientists, **the data provider is responsible for providing high-quality data.** Low-quality data is data that does not meet the requirements for its format, completeness, reliability, relevance and other characteristics necessary for the correct operation of AI [153].
- According to a study by Russian law firm Intellect, **the developer selects the necessary data for training, correcting the results and evaluating the accuracy of the data during training**. It is the developer who informs AI what is reliable and what is not — the criteria of 'reliability' are usually checked for validity and compliance with legal principles [154].
- **The customer determines for what purposes the AI will be used** and what data is needed to achieve this. It also monitors and analyzes the results of the system for the issuance of false information.

Recommendations for data providers:

1. **Information about the data should be disclosed**, for example, about its origin, methods of collection, and known limitations and distortions should be noted.

2. **Ensure that the data is updated regularly.**
3. **Warn contractors** about changes in the datasets you supply.

Recommendations for developers:

1. **Pay attention to the data sources** for training and taking into account their specifics.
2. **The data provided should be subject to preliminary analysis.** If problems are found, developers need to notify the data providers.
3. **It is recommended to test the model.** This will help to identify unreliable data.

Recommendations for customers ordering models:

1. **Ensure that the data is checked for compliance with the assigned tasks**. Also participate in the reconciliation of the dataset.
2. **Monitor the operation of the system.** Distortions of information identified in a timely fashion can be eliminated by further training the model.

Research on the issue:

According to an American study on 'Ways to ensure data quality for machine learning'[155], the term **'qualitative data'** refers to purified data containing all the attributes on which model learning depends. This study also provides 4 characteristics of qualitative data for training ML models:

- **Relevancy** — the dataset should only contain features that provide meaningful information for the model.
- **Consistency** — similar examples should have similar labels, ensuring dataset uniformity.
- **Uniformity** — the values of all attributes must be comparable for all data.
- **Comprehensiveness** — the dataset must contain a sufficient number of parameters or features so that there are no borderline cases left uncovered.

What do the experts think?

Ivan Oseledets,
General Director of the AIRI Institute

“If a person specifically wants to lead the model to ‘behave badly’, then they are responsible for this. If the model itself immediately starts talking nonsense, then, of course, the question must be posed to the developers as to why they have not checked it. It seems to me that we need to go in the direction of creating experimental legal regimes, giving the right to make mistakes. The main problem is that AI is now a gray area, and no one wants to start using it on a large scale, because “what if?”.

Anna Meshcheryakova,
CEO of the “Third Opinion”

“We use open datasets at the research stage. We work with data published in Russia and interact with foreign colleagues. Our own scientific activities and cooperation with medical and technical universities in Russia and abroad allow us to obtain high-quality datasets for research purposes. But at the learning stage, we rarely use open datasets — we have our own requirements for classifiers and markup[156].”

Denis Dimitrov,
Managing Director of Data Research,
Sber AI

“The fight against false information in the training of artificial intelligence models is a complex process that requires work in several directions: the use of high—quality data sources, filtering and cleaning data, manual data verification, the development of fact-checking models and processing feedback from users. In addition, one of the ways to combat unreliable answers and model hallucinations is retrieval — further training of the model in order to use external knowledge and data bases (for example, the Internet).”

17 The problem of spreading malicious or misleading information using AI: how to mitigate this?

Answer:

To prevent the spread of malicious or false information using AI, it is necessary that everyone involved in the creation and use of this technology, from developers to users, be responsible for their own input within the ecosystem, taking into account legal and ethical standards.

Justification:

- **The issue of the spread of malicious and misleading information through AI is more ethical than technological in nature.** At the same time, it makes sense to put possible protective measures in place on a development level, in order to prevent the use of AI for unintended purposes.
- A group of European scientists from MediaFutures argue that **cooperation between platforms, governments and civil society contributes to effective content moderation**, the dissemination of fact-based information and compliance with the law [157].
- **Human-oriented and humanistic functioning of AI systems includes their responsible development and correct use.** These two are vital components in creating a safe, reliable and ethical AI that can benefit people.

Research recommendations:

1. **Encourage initiatives by development companies to verify neural networks created as well as undertake voluntary certification.** Development companies should be aware of the intended application for AI and any potential liability for non-compliance with established requirements. At the same time, it is important to consider that due to the work of generative AI, the information generated may not meet the expectations of users.
2. **Raise public awareness on this issue.** Launching initiatives to train people's critical thinking and digital literacy skills will allow people to recognize and filter potentially false information.

Recommendations for developers:

1. **It is important that filters and barriers, such as censors, should be used.** This will prevent the creation of clearly toxic content.
2. **It is recommended to use benchmarks to check generative neural networks** for the correctness of generations.
3. **When there is a technical possibility to do so, the source of the information should be indicated.** Users can then in turn decide for themselves whether to trust this information or not.
4. **Eliminate model errors identified during use** related to data inaccuracies and actual distortions.

Recommendations for users:

1. **Use AI only in accordance with legal requirements and platform rules.** Unfair use of AI tools (for example, to create offensive deepfakes) is considered unethical and may lead to negative consequences for the user (account blocking).
2. **Keep in mind that AI systems can sometimes 'hallucinate' and output false information.** Critically analyze the information and use tools to verify the authenticity of output generated.

Research on the issue:

1. In June 2024, UNESCO published **a report on the risks of using generative AI for applications of Holocaust remembrance**[158].
   This report focuses on the fact that AI may 'hallucinate' and produce fictional facts. For example, chatbots very often distort information about the number of victims of the Holocaust. Also, systems cannot always correctly evaluate distorted information that is only partially false (for example, that all Nazi concentration camps had gas chambers) or opinions (for example, that gas poisoning was the worst kind of mass murder during the Holocaust).
   As a solution, UNESCO suggests that developers use a wide range of data for training, consult with stakeholders on sensitive topics, and bring their risk monitoring and assessment systems in line with ethical principles. In turn, users should understand the limitations of AI technology and independently verify the authenticity of the content.
2. According to a survey conducted by the Russian Public Opinion Research Center, over the past year **society's demands for clear distinction and labeling of products created using artificial intelligence have grown stronger**[159].
   Since 2023, the share of people in favour of the mandatory labeling of AI output has increased from 69% to 73%. At the same time, the share of people opposing this measure has decreased from 23% to 17%, while the share of people who are categorically opposed has halved (from 14% to 7%).
   This trend indicates an increased awareness among Russians of the importance of preventing the spread of malicious or misleading information using AI.

What do the experts think?

Daniil Gavrilov,
Head of the AI Research Laboratory at T-Bank AI Research

“Automatic detection of fakes and malicious information is a difficult task for the market, because everyone has a different understanding of what can and cannot be considered acceptable accordingly. There are methods that can reduce the number of unsafe texts, but they do not guarantee full protection against vulnerabilities.
One area of focus that can help solve the problem is the development of model interpretability methods. They allow you to get an answer to the question of: “Why does artificial intelligence offer a specific solution in a given situation?” This gives an opportunity to better understand the internal processes of AI and prevent undesirable results. This field began to develop rapidly after large language models became available to a larger audience. At T-Bank, we pay special attention to this through scientific research.”

Oleg Yangalichin,
Executive Director of Data Research, Sber

“We use a multi-level approach to prevent the spread of malicious or misleading information using AI. This includes both the development and implementation of algorithms for automatic detection and blocking of disinformation, as well as regular validation by checking models for vulnerabilities that can lead to unsafe output generation. Another important element is the user feedback system, which allows them to respond quickly to any potential threats.”

18 Is it ethical for algorithms to offer the user goods and services that do not correspond to their usual preferences?

Answer:

Yes, if a algorithm works without bias and offers a variety of goods and services to all users, it can be ethical.

Justification:

- A group of researchers from India and the UK believe that **offering the user goods and services that do not correspond to the user's usual choice helps to avoid the formation of an information bubble.** This approach broadens horizons and does not limit the user to their usual preferences [160].
- In a study conducted as part of the Implementation of the Code of Ethics for AI says the use of user data for **the operation of recommendation services is legitimate** if the developer complies with legislation on personal data and the requirements of other regulations [161].
- Researchers from Rutgers University argue that **if a company has data on user preferences, it is important to take them into account when making offers.** Ignoring this data may be perceived as disrespectful to user [162].

Recommendations for developers:

1. **Develop transparent algorithms.** The algorithms used by the AI to make suggestions should be transparent and understandable to the user.
2. **Take into account the interests of the user.** An offer can be considered ethical if it is relevant to the user to some extent. For example, if a user is interested in healthy eating, it may be unethical to offer them sweets and fast food.
3. **It is recommended not to use algorithms to recommend extremely sensitive categories of goods and services.** For example, adult goods and services, those of a religious or ritual nature; or those which may contribute to inciting conflicts or inter-ethnic tensions.
4. **Provide a choice.** The user should be able to set preferences and refuse to receive those that they are not interested in.

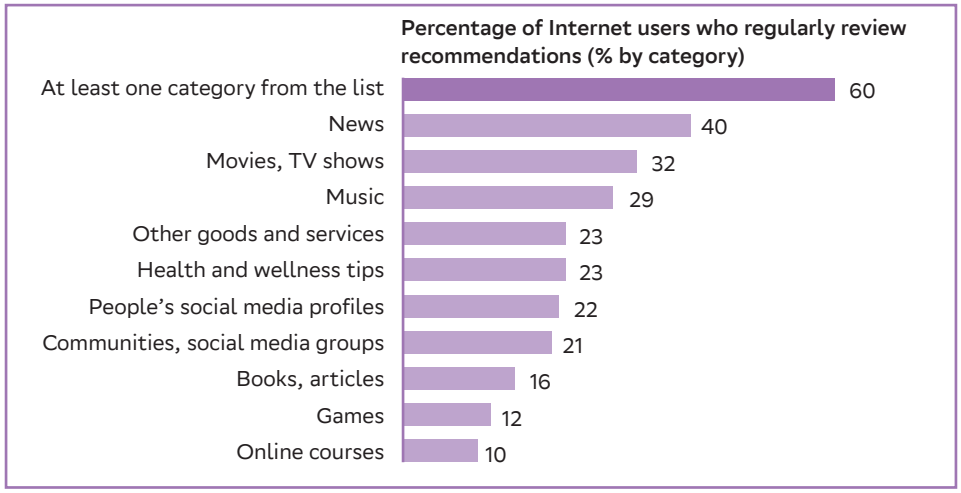
5. **Inform people about the technology in use and the reasons for the offers.** Explaining the reasons why a particular product has been offered to a user builds trust. For example, if there is a promotion, or a new offering that's worth trying.

6. **Analyze feedback.** Regularly collect and analyze feedback from users to improve algorithms and suggestions.

Research on the issue:

1. In 2023, the Institute for Statistical Studies and Economics of Knowledge at the Higher School of Economics in Moscow conducted a survey of people aged 14 years and older. The results showed that **the majority (60%) of Russian internet users (who go online at least occasionally) will often or almost always view the recommendations given by digital services**. The most interesting topics are news (viewed by 40% of internet users surveyed) and entertainment resources (movies and TV series — 32%, music — 29%)[163].

2. According to a McKinsey study from 2021, **71% of consumers expect companies to provide personalized interaction**, and 76% are disappointed when this does not happen[164].
Moreover, personalization improves productivity and customer service. Companies that grow faster earn 40% more revenue from personalization than their slower-growing counterparts. And companies that have succeeded with personalization receive 40% more revenue from this activity than the average market player.



Source: McKinsey[164]

What do the experts think?

Alexey Byrdin,
General Director of
the Internet Video Association

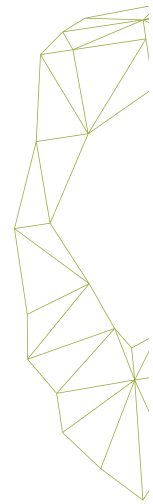
“Hybrid recommendation systems that use collaborative filtering will in any case ‘break through’ a user’s ‘information bubble’. This makes it possible to broaden a user’s horizons and introduce them to new things, including products or services that have appeared, which is perfectly ethical. But in order to avoid irritating or upsetting users, it is important to avoid making recommendations of certain (overly niche) categories of items that are too far removed from their announced circle of interests.”

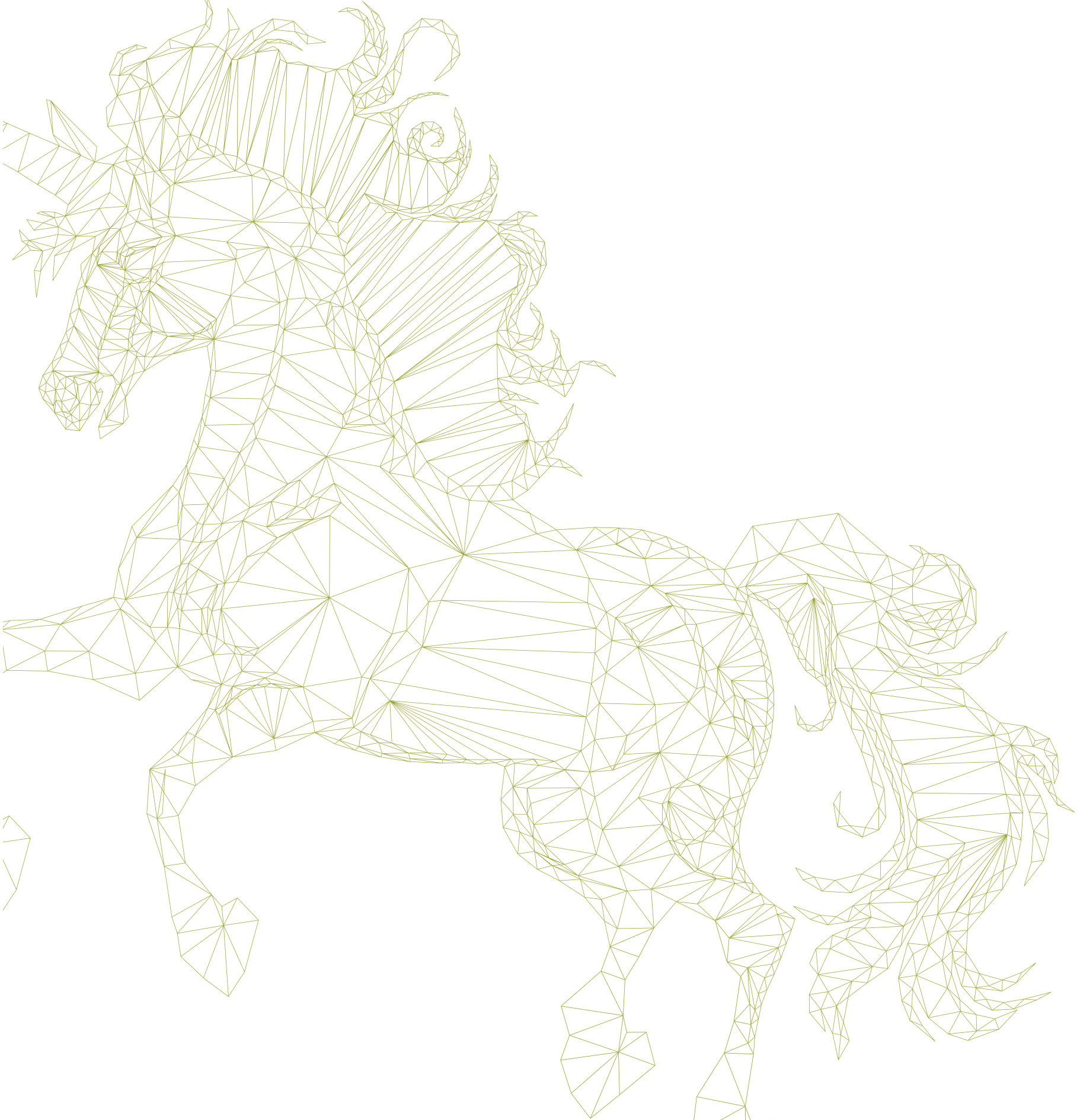
Andrey Zimovnov,
ML Director, VK AI

“Our recommendation systems process tens of billions of user signals every day, from views and listens to likes, shares and comments. This allows us to make recommendations in our services more accurate and relevant. At the same time, it is important for users to show not only the content that they are used to watching or listening to. This avoids the formation of a ‘dopamine loop’ and/or an ‘information bubble’. To do this, we have developed the Discovery mechanism. It offers users not only what they are already watching, but also new authors or even whole new topics.”

Chapter 04

Generative AI





19 Is it possible to trust information obtained with the help of generative AI and AI-based search engines?

Answer:

Both search engines and generative AI are different, not mutually exclusive ways of obtaining information from the accumulated knowledge. Together, these two technologies help the user to get an up-to-date and understandable answer to their question: the search finds relevant documents, and the generative AI formulates the answer from what it finds. Any information, including information obtained with the help of generative AI. AI should check and aim to find the primary sources.

Justification:

- **Russian scientists emphasize that most search engines operate on the basis of AI technologies** (primarily for ranking results). AI allows you to search pages not only for specific words, but also by meaning, then personalize the output, generate hints, and so on[165].
- **Both search engines and generative AI can give false answers**, because they learn from data sources from the internet, which may contain errors, so it is important to refer to primary sources or consult with experts regarding the relevance of information.
- **Generative AI can 'hallucinate', which means outputting imaginary facts.** This happens in a situation where generative AI uses its own knowledge to respond. If there is no direct answer to the user's question in the training dataset, the AI tries to deduce it based on general patterns.
- **Using machine learning systems to classify and categorize large amounts of data allows you to speed up processing and improve the accuracy of search results.** AI algorithms are able to analyze and interpret texts, images and videos, so that manual processing of these types of data takes less time.
- The responses of generative AI may be limited by the amount of data that was used to train the models. Therefore, if a model is developed, for example, based on texts produced before 2023, it will not be able to comment on the news of 2024, or its answers may be outdated.

Recommendations for developers:

1. It is important to make the rules and principles of search engines and services based on generative AI transparent and public. This will increase user confidence in the technology.
2. It is necessary to warn users about possible errors and limitations of these technologies, as well as inform them about the sources of responses.
3. AI-based services cannot be used to impose a certain point of view or to persuade users to make a decision. Search engines and services based on generative AI do not create barriers to obtaining information, with the exception of illegal, malicious or life-threatening content.

Recommendations for users:

1. It is necessary to be critical of the information received from search engines and from generative AI, and to double-check such information using reliable sources or primary sources.
2. It should be remembered that the responsibility for spreading false information received from the chatbot and for making decisions based on such information lies with the user.
3. It is important to pay attention to the conditions of use of AI systems, where the developer can warn about certain risks associated with the operation of the system in terms of information processing.

Research on the issue:

1. **Researchers at Voronezh State Technical University compared the results from the Yandex and Google** search engines with the AI-based **ChatGPT system**. The analysis showed that when responding to user queries, Yandex and Google provide the most relevant links containing all the keywords from the query. Meanwhile, ChatGPT can immediately provide a structured response that covers all important aspects of the question[166].
   However, ChatGPT has a drawback: the system does not provide links to information sources, which makes it difficult to verify authenticity. In addition, sometimes ChatGPT gives incorrect answers. For example, when asked about the name of the first human cervical vertebra, the system gave the wrong answer, while Yandex and Google provided brief and correct information, as well as links to data verification resources.
2. A study conducted by scientists from the University of Washington has shown that **systems running on generative neural networks can malfunction and generate absurd results for no reason**[167].
   As an example, the researchers turned to the Perplexity AI and the Arc search engine with a request to provide information about a non-existent theory called ‘Jevin’s theory of social echoes’. In response, AI proposed a concept and even provided links to non-existent sources.

The Commission's approach to the implementation of the Code of Ethics in AI:

In 2024, representatives of the business community signed the Declaration on Responsible Generative AI. It contains a number of recommendations and standards of behavior in generative AI technologies for developers, users, representatives of the academic community, as well as everyone who creates, implements and uses generative AI technologies. The document emphasizes that generative AI is one of a number of possible tools that can be use.

Practices:

Google has published search rules regarding AI-generated content.
The company writes that they prefer unique high-quality content that meets E-E-A-T standards (Experience, Expertise, Authoritativeness, and Trustworthiness). The company also says that about ten years ago it faced the problem of rapid growth in the volume of content created by people. Blocking all such content would be unwise. Therefore, the company decided to improve its systems so that high-quality content is given an advantage. Thanks to ranking systems and determining of useful content, users are provided with materials created primarily for people, rather than in order to improve the rating[168].

What do the experts think?

Marina Rossinskaya,
Chief Operating Officer of Yandex Search

"When we offer the user a search engine response created using generative neural networks, we always inform them this is the case. It is also important that such responses always contain links to the sources on the basis of which the response was generated. This allows the user to go to the sites and double-check the information or find out additional facts."

Daria Chirva,
Researcher at the Center for Strong Artificial Intelligence in Industry, lecturer at the Institute for International Development and Partnership at ITMO University

"A person is responsible for any statement. The generation of texts using large language models does not relieve them of this responsibility due to the fact that no one and nothing else can handle it yet. Checking any fact, value statement, etc. obtained with the help of AI tools is a necessary element of their proper use."

20 Is it ethical to synthesize human speech using AI?

Answer:

It is ethical to synthesize the speech of an existing person in some cases, for example, to create works of art, but only if there is an consistent consent of the person whose voice will be used in speech generation services.

Justification:

- Fliki, a major developer of AI solutions for creating video content, indicates that **transparency and consent are of paramount importance in the ethical use of AI voice cloning**. Creators should seek explicit consent when using cloned voices, especially in scenarios where the cloned voice is used for commercial or public purposes. Consent ensures that people have control over the use of their voice, and prevents unauthorized or unethical voice cloning [169].
- In combination with audio devices, AI generation technologies can become indispensable assistants **for those who have lost the ability to speak, or for blind people.**
- Speech synthesis is used not only when **creating audiobooks or podcasts, which allows you to 're-sound' content depending on the listener's preferences, but also in everyday life. For example, in maps and navigators — when voicing a route — or for voice assistants, which is also an ethical** use case provided there is consent.
- It is also **ethical to clone voiceswith the help of AI for simultaneous translation**, which is no different from the use of stand-in voices, which is a widely used practice in the creation of audio and video works.

Recommendations for developers:

1. It is important to **explain to the owner of the voice that is used to train the model the features of speech synthesis technology and note that the nature of texts that will be voiced is unknown in advance** and will definitely differ from those voiced for the training dataset.
2. Always try to regulate the issue of voice synthesis by contracts.

3. **The use of the voices of public people** contained in publicly available sources is possible within the framework established by law (for example, for parodies). If there are no such restrictions in the law, the use of voice should not humiliate the honor and dignity of its bearer or be used for illegal purposes or in violation of accepted moral norms.
4. In agreements on the use of technology, it is **advisable to reserve the right to revoke access to the service** in order to block users who create and/or distribute illegal content.
5. Confidentiality of the received data should be ensured and their leakage should be prevented.

Recommendations for users:

1. When disseminating content created by generative AI using someone else's voice, it is **important to attach a clear message that the content was generated by another person using AI.**
2. When distributing content with elements of another person's personal data, **the relevant consent should be requested.**
3. **When using voice generation services**, do not use it for illegal purposes or in violation of accepted moral norms.

Research on the issue:

1. According to Russian scientists from the Putilin Belgorod Law Institute of Ministry of the Interior of Russia, **one of the main problems of voice cloning using AI is the possibility of wrongdoers to use the technology for fraud or to spread disinformation**[170]. Cloned voices can be used to deceive, manipulate, or steal personal data, leading to serious ethical and legal violations. For example, the use of a person's voice and face can allow criminals to spoof their photo and video images to illegally obtain loans, change the ownership of real estate, or discredit any legal entity or individual.
2. Neuroscientists from Switzerland have found that **in the auditory cortex of the human brain there are mechanisms that allow you to identify a voice created with the help of AI**. For the experiment, the scientists synthesized the voice from a recording of a real person and recorded the brain activity of 25 listeners using a functional MRI scan[171].

Practices:

1. According to AI video and audio content solution developer Fliki — **3 seconds of audio is enough to create a voice clone that has 85% coincidence with the original**[172].

2. In Maryland, USA, a physical education teacher generated the voice of a school principal and distributed their allegedly racist and anti-Semitic statements to teachers. The incident occurred after a professional conflict between teachers[173].

What do the experts think?

Alexander Krainov,
Director of Artificial Intelligence Development, Yandex LLC

"Of course, the conditions needed for the use of voice, the use of audio recordings, the process of voice transmission and models based on speech synthesis to third parties are set out in the law. But, as practice shows, this is not enough. It is necessary to inform the speaker as fully as possible about all possible ways that their voice will be used. The speaker's decision to provide their voice for speech synthesis service must be made fully informed."

Alexey Parfun,
CEO of Agenda media group, co-founder of Reface Technologies, Vice President of ACAR

"Using voice clone technologies is an understandable and very useful tool that, combined with a number of other technologies such as lip sync, allows media content producers to significantly reduce costs and increase production speed. Like many other tools, it can turn into a dangerous weapon in the hands of wrongdoers, so you should use labeling and hardware control on the side of social media in order to prevent fraud."

21 Is it ethical to use generative AI in art and design?

Answer:

It is ethical to use AI at any stage of the creative process, but subject to compliance with ethical and legal standards by both the developer of generative AI and the user.

Justification:

- **AI embodies a person's intention**, therefore, the responsibility for the ethical use of AI lies with the individual (as the bearer of meaning) who sets the AI task. At the same time, one should pay attention to the results of AI and remember that AI can 'hallucinate' and produce technical errors.
- An article by the Moscow School of Contemporary Art **expresses the view that AI greatly simplifies the process of creating works of art for artists and photographers. It helps in choosing images, developing ideas, presenting projects and finding inspiration**[174].
- AI **can also be used to recognize fake works with 90% accuracy**, which will help restore the integrity of art history[175].
- Researchers at the Katanov Khakassian State University suggest considering **AI as a tool for creating art on an equal basis with other methods (computer software, paint brush, etc.)**[176].
- Researchers at the Institute of Electrical and Electronics Engineers (IEEE) believe that **AI democratizes creativity, making it accessible to a wide range of people, including those who do not possess special skills, as well as for people with disabilities**[177].

Recommendations for users:

1. **Review the user agreements of generative AI platforms** to understand how the rights to the created content are distributed.
2. **Distributors of AI content that mimics real events should explicitly indicate its origin.** For artistic works, labeling is not essential, unless it is otherwise required by law.
3. **It is necessary to weed out unethical or illegal generated content and report it to the developer** (in support), since despite the limitations of the platforms, there is still a possibility of this outcome.

4. **You should keep information about the parameters of the AI system, the datasets used, and your contribution to content generation.** This can help to substantiate originality and claim the rights to the created works.

Recommendations for developers:

1. **You should take into account copyright rules** when teaching a generative model or creating your own datasets.
2. **The generated content should be moderated and restrictions should be placed** on sensitive or illegal topics.
3. **A feedback form should be provided for the user** to be able to take into account concerns when moderating the generated content.

Practices:

In April 2023, the finalist of the World Photography Organization's Sony World Photography Awards was announced, with the award going to photographer Boris Eldagsen for his work "The Electrician". However, he did not accept the award, explaining that **his photo was generated using a neural network**.

In response, The World Photography Organization cut ties with the artist, declaring his intentions dishonest while recognizing that Boris raised an extremely pressing issue about the need to differentiate and redefine many categories and forms of art.

Research on the issue:

A study by Market Research showed that in 2022 **the global market for generative AI** (visual arts, music and literature) in art was estimated at 212 million US dollars, which is expected to **reach 5.84 billion dollars by 2032**, demonstrating an average annual growth rate of 40.5% during the forecast period[178].

UNESCO's approach:

UNESCO recommends that States promote AI education and digital learning for artists and creative professionals to assess the suitability of AI technologies for use in their profession[179], and also promote the development and implementation of appropriate AI technologies, since AI technologies are used to create, produce, distribute, broadcast and consume various cultural goods and services, and given the importance of preservation of cultural heritage, diversity and freedom of creativity.

What do the experts think?

Anna Kulik,
Marketing Director of 'Inferit'

"It is pointless to compare AI art with traditional art, as it is to compare different creative forms and genres. AI is an assistant to the creator, a tool. It is the creator who is responsible to society for observing ethical standards in the working process, and also for the end result. Generative AI tools allow everyone today, regardless of their level of skill and knowledge, to express their unique vision of the world. To draw with a word, voice or thought, or to create musical works without knowing musical notation is a gift to humanity."

Ivan Shumeyko,
Art director of 'Inferit'

"AI in art is a fascinating tool that can offer a non-standard view and make certain technical tasks easier. But a person always puts their own essence into a work. Only personal experiences, emotions and the inner world of the creator can truly touch the viewer and evoke a response in their heart. AI can be a virtuoso assistant, but without a human spark. It will never create a masterpiece that will make us laugh or cry, empathize or dream."

22 Is it ethical not to indicate that content has been generated with use of AI?

Answer:

More and more synthesized content is being created now, and technology is often used to refine materials made by humans. Therefore, there is no definite answer to the question. If images, text, audio or video generated by AI can mislead people about their origin, especially when it is important, it is unethical to use such content without explicit labeling. It is always necessary to take into account the context and purpose of using AI.

Justification:

- According to the Oxford scientist, it is important to consider different types of labeling based on the situation[180].
  Visible markings clearly noticeable to users (for example, the text "Getty Images" on pictures).
  Invisible markings containing technical signals embedded into the content.
  Both types of watermarks — known as 'direct' and 'indirect' disclosure — are important to ensure transparency.
- Scientists from the Massachusetts Institute of Technology (MIT) believe that as generative AI systems are increasingly able to create high-quality media, **visible and invisible labeling of AI-generated content offers potential protection against deception and mix-ups between original and AI content**[181].
- Researchers from South Ural State University believe that **labeling will increase confidence in both creators and owners of generative AI systems, as well as the generated work itself**[182].
- **The distribution of 'fakes' — content indistinguishable from the real thing — without appropriate labeling can be perceived as manipulation** and negatively impact upon reputation.

Recommendations for developers:

1. **In cases where the way of creating content using AI is fixed, you can use invisible labeling, which does not affect the appearance and quality of the content** and allows the user to use it freely. Such labeling will protect the rights of the user and the developer if any violations are detected by supervisory authorities.

2. **In some areas, it is possible to provide the visible labeling** so that the user of the service understands that what they see is an AI generated content.

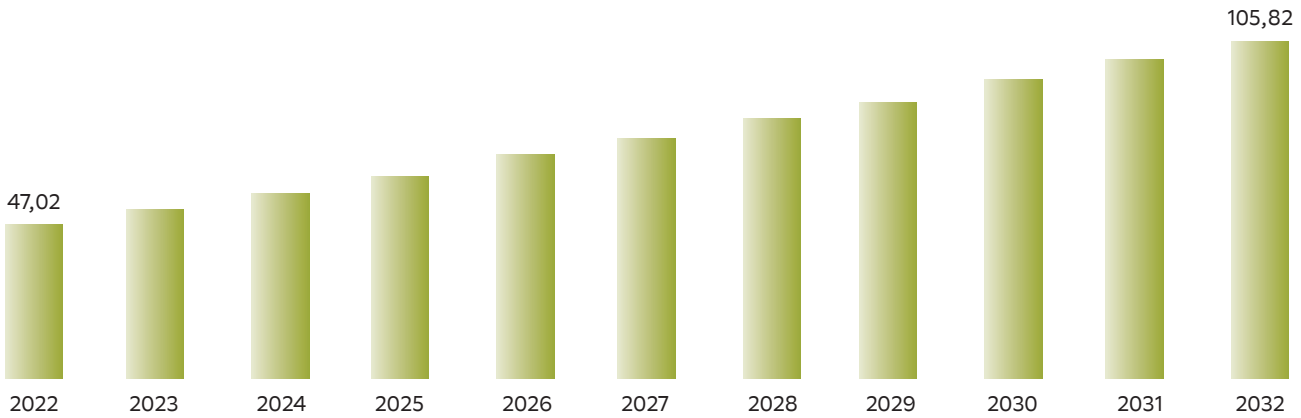
Recommendations for users:

1. **Depending on the context, it is important to indicate that the generated content was created using AI when distributing the information, not to mislead other users about the personal authorship of this content and not undermine the credibility of their publications.**

2. **Treat markings set by the developer responsibly** and do not try to circumvent or hide them.

Research on the issue:

1. Based on a Business Research Insights study, the size of the global digital watermark technology market in 2022 was US$47.02 million, while it is forecast to reach US$105.82 million by 2032, representing a CAGR of 8.45% over the forecast period [183].



Source: Business Research Insights [183]

The labeling technology market has experienced significant growth in recent years, driven by the growing need for secure and authentic digital content. Concerns about intellectual property theft, forgery, and unauthorized use of content have fueled demand for reliable watermark solutions.

2. According to a Stanford University scientist, **label manipulation is a concern**. For example, invisible watermarks are often promoted as the leading solution for labeling AI-generated content, with embedded markings much easier to manipulate in text than in audiovisual content. The AI content labeling policy should be specific about what kind of content invisible watermarks are useful for, since a particular disclosure solution used for images is not necessarily useful for text.

Practices:

In image-based systems, watermarks function by adding subtle noise to an image (for example, slightly changing every seventh pixel) to create a cryptographic marker[184]. However, **text watermarks are more difficult to create because there are limited ways to alter text without changing its meaning**. For example, a recent case: the company Genius.com filed a lawsuit against Google to remove song lyrics from its website. To prove their point, in one of the song lyric texts on their website, they switched some curly apostrophes for straight ones. The sequence created translated to "red-handed" in Morse code. According to the lawsuit, this sequence duly appeared on the Google platform, indicating that it had been copied from Genius.com[185].

What do the experts think?

Anna Abramova,
Director of the AI Center at MGIMO University

"Standardization in the labeling of generated content will increase transparency for artificial intelligence technologies. The development of national standards in this area will serve as the basis for the formulation of proposals for international cooperation."

Sam Altman,
Chief Executive Officer, Open AI

"I do want to flag something else that I think is underexplored, which is the idea not just of watermarking generated content, but authenticating non-generated content. Celebrities or politicians be able to "cryptographically sign" messages to prove that they actually produced them. That seems to me like a reasonably likely part of the future for certain kinds of messages and I think we should talk more about that[186]."

23 Will generative AI affect standards in beauty and fashion?

Answer:

The risk of AI influencing beauty standards is quite high, since AI filters using machine learning algorithms can create an unrealistic image of a user's appearance. They smooth the skin, change the shape and size of the face, apply virtual makeup creating an idealized appearance of an individual.

Justification:

- American scientists noted in a study say **that due to the pursuit of a perfect appearance on social networks, a new disease has arisen, called 'social media dysmorphia'**. This is a mental disorder in which a person suffers from excessive anxiety about their appearance.1 AI filters and AI models can exacerbate the spread of this disorder[187].

Recommendations for developers:

1. **Try to take into account the uniqueness of each person: everyone has a unique appearance and style, and template images often do not take this into account.**
2. **Develop critical thinking and use image analysis** when using social media to recognize manipulative standards of beauty created by AI.
3. **Healthy beauty ideals that are based on a healthy lifestyle** and self-esteem should be supported.

Practices:

Fashion company Levi's generated a model using the service LaLaLand.ai, a digital studio that creates models for fashion companies using AI. **The use of virtual models also increases the risk of job losses among real models.** Digital models can offer a wide range of posing options, imitating the behavior of a living person, and being able to do so without experiencing fatigue. In addition, as the founder of LaLaLand.ai Michael Musandou emphasized, businesses can save not only on models, but also on makeup artists, photographers and other personnel involved in shoots of[188].

In the summer of 2024, **the world's first beauty contest for non-existent models** took place. "Miss AI", the digital equivalent of the well-known Miss World contest, was held via the online Fanvue platform and organized by World AI Creator Awards (WAICA)[189]. Only generated images were accepted for the competition — with photos of real people strictly filtered out. The Miss AI Jury included Miss Great Britain, marketing experts and creators of popular AI products, and the total prize fund stood at 20,000 dollars. The winner of the contest was Moroccan AI blogger Kenza Leyli.

Research on the issue:

Plastic surgeons at the Albacete University Hospital in Spain emphasize in their article that we need to be careful what AI already finds out about us as of right now[190]. **It is important to eliminate biases and misunderstandings for AI systems**, especially those that may perpetuate harmful stereotypes or unrealistic standards of beauty. This paves the way for further research to develop more inclusive and diverse AI models that better reflect the diversity and complexity of human beauty.

What do the experts think?

Dr. Kerry McInerney,
Researcher at the Leverhulme Center for the Study of the Future of Intelligence at the University of Cambridge

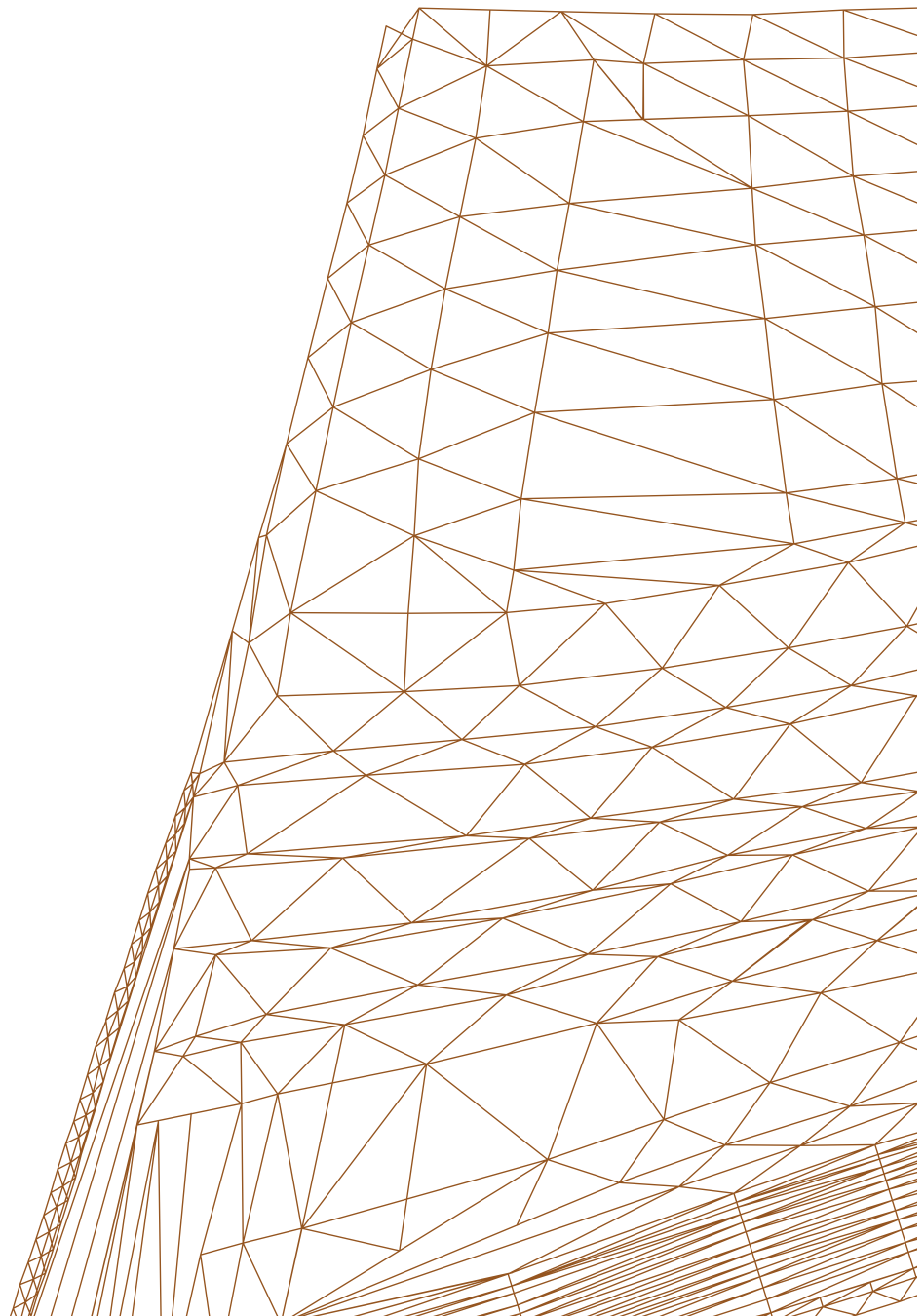
"Most of the models that made it onto the list of contenders for the title of "Miss AI" were light-skinned and slender, making them not dissimilar from the real life models we are familiar with. AI tools are designed to replicate and scale the world's existing beauty standards, rather than challenge or change them[191]."

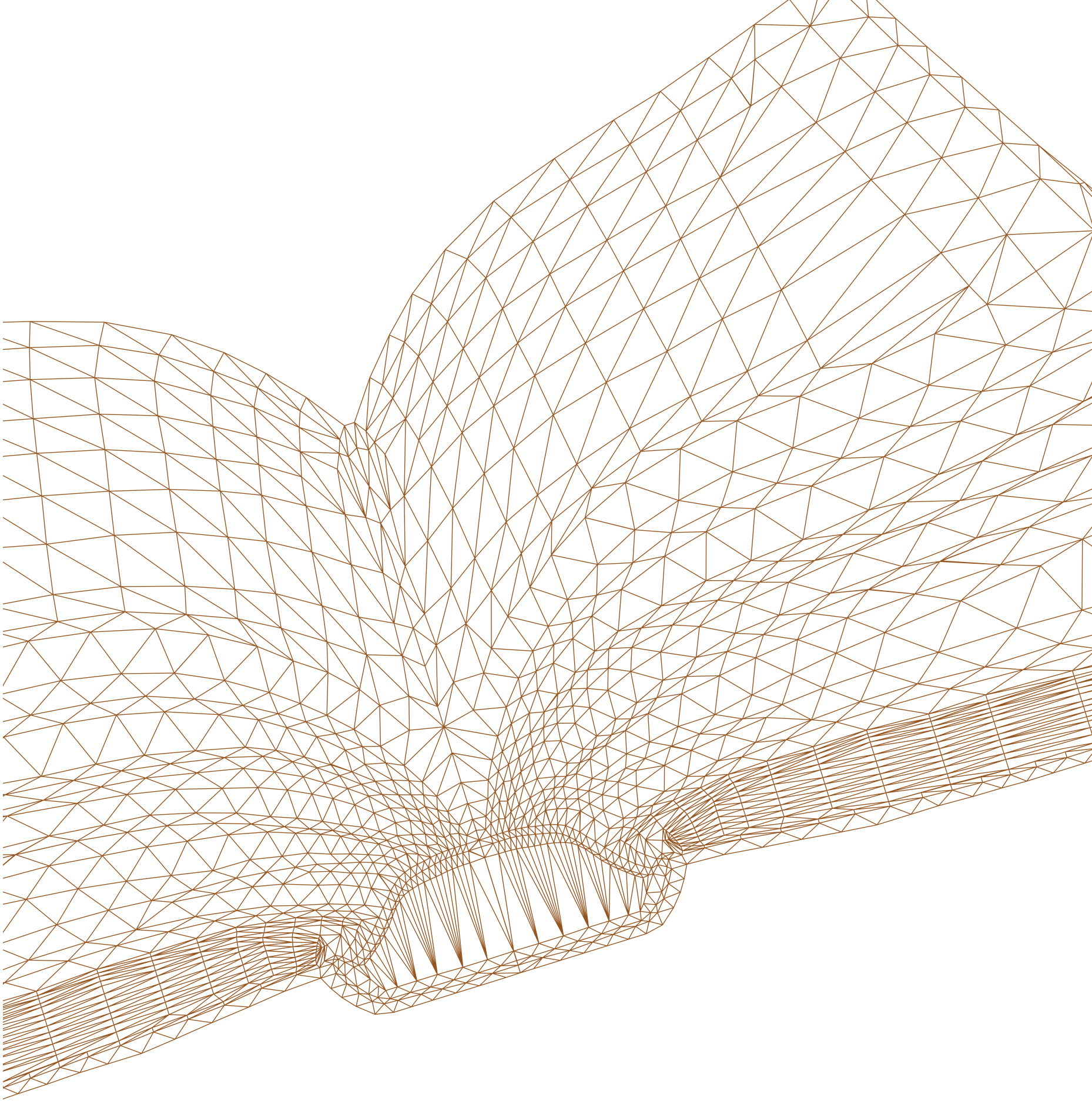
Jennifer Levine,
American certified plastic surgeon

"We are faced with heavily AI-edited images that people are starting to see as beauty standards. I think that in the future a great many people will criticize these images, which will make it possible not to use such strong transformations[192]."

Chapter 05

AI in education





24 Is it acceptable for students and teachers to use AI in the education process?

Answer:

The use of AI technologies in the educational process is acceptable for both teachers and students, provided the requirements of legislation, age restrictions, the specifics of a particular area of educational nature and the internal rules of a particular organization are taken into account.

Justification:

- According to a joint study by Staffordshire University and the Georgia Institute of Technology, **modern AI systems can reduce the time required to complete most routine activities**, freeing up time for creative tasks[193].
- **Automation of feedback and the marking of assignments** eliminates the human factor and bias, providing students with timely and objective feedback, while simplifying the assessment procedure for teachers.
- Companies providing services for preparing for various exams note that AI-based adaptive learning systems take into account the capabilities, interests, needs and individual characteristics of the student[194].
- A group of American programmers developing AI for education believe that the use **of AI technologies contributes to the development of hybrid formats** and simplifies access to educational materials regardless of time and location[195].
- A group of researchers from India and the UAE recall the importance **of human contact in the educational process**. It promotes the development of social ties, creative and intellectual potential, so therefore replacing it entirely with algorithms puts the establishing of responsibility and motivation for students to realize their potential at risk[196].

Recommendations for an educational organization:

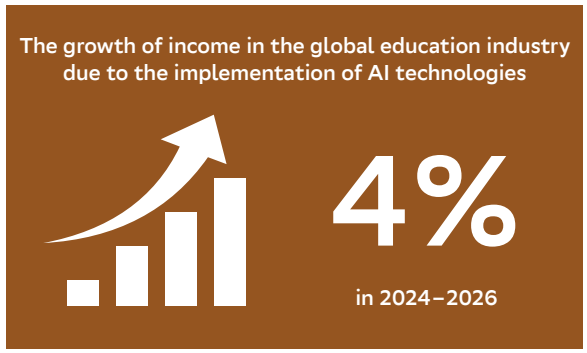
1. **AI technologies should be introduced into the educational process in moderation, not forgetting the importance of social connections.** It is recommended to find a balance between the use of technology and the preservation of traditional teaching methods that promote the development of emotional intelligence and interpersonal communication skills.

2. **Create educational programs for teachers and students on how to work correctly with AI tools.** This will help teachers effectively integrate these tools into the learning process, and students will be able to learn how to use these tools to solve various tasks.
3. **Define clear educational goals for using AI.** For example, AI can be used to personalize student learning or to automate routine tasks for teachers.
4. **Consult with the expert community and refer to the results of scientific research.** This will allow you to select the appropriate tools and technologies, and also develop a training and support plans for teachers and students.
5. **Confidentiality should be respected.** It is necessary to provide a legal basis for the collection, use and processing of personal data, including social and ethical considerations.

Research on the issue:

1. According to McKinsey's annual research on the state of the AI technology market for 2023[197], industries closely related to knowledge are likely to undergo the greatest changes. At the same time, **these industries can also receive significant benefits**.

2. In 2024, UNESCO published its **'Guide to the Use of GenAI in Education and Research'**[198]. According to UNESCO, despite growing attention to the development of thinking and creativity, the importance of basic skills for the psychological development of children and the building of skills among students is beyond doubt. These fundamental skills include listening, pronunciation, and writing in a native or foreign language, as well as the basics of numeracy, drawing, and programming. The "exercise and practice" approach should not be considered as an outdated pedagogical method; instead, it should be actively used and modernized using generative AI technologies. If ethical and pedagogical principles are followed, generative AI tools can become individual trainers for practice through independent learning.



Source: McKinsey[197]

What do the experts think?

Eduard Galazhinsky,
Rector, Tomsk State National Research University

"One of the main and absolutely new challenges for higher education is the need to learn how to function adequately as soon as possible, in an environment where content generation using AI technologies becomes accessible to every researcher, university teacher, student and graduate. This challenge clearly demonstrates the ambivalent nature of any technology. At first, as a rule, it seems to be an undoubted benefit that makes people's lives easier, simpler and more pleasant, but very soon its negative features begin to show through. Universities are now actively engaged in this task."

Elena Bryzgalina,
Head of the Education Philosophy department at the Lomonosov Moscow State University (MSU) Faculty of Philosophy, Moscow State Head of the Bioethics Master's Program

"In the current conditions, it is necessary to train students to interact with AI effectively, as a tool for solving working tasks while taking into account the standards of academic ethics. With the increasing integration of AI into the educational process and science, clarification of the conditions for the ethically acceptable use of AI should be accelerated. By-laws of educational and scientific institutions or methodological regulations for certain types of educational activities (studying a specific discipline, conducting practice, etc.) can serve as the documents that establish ethical boundaries."

25 Is it ethical for a teacher teach a subject using digital imitation without an actual presence in the classroom?

Answer:

No, it cannot be considered ethical. Digital imitation cannot be a full-fledged substitute for an 'in-the-flesh' teacher in a classroom. The use of such technology is allowed only in certain cases and under certain conditions.

Justification:

- **As an experiment it can be used in the classroom, it is preferable to use a digital imitation of the teacher only in certain situations**, of online education outside classroom hours. Mutual respect, understanding of socially acceptable and ethically correct behavior, and skills of working with meanings and values can be formed only through interpersonal communication.
- If using a digital imitation, the teacher should be concerned about **preserving the social and communication skills of students** and preventing devaluation of human interaction.

Recommendations for teachers:

1. Discuss the advantages and disadvantages of using this tool with students and experts in this field. This way, you will be able to prevent potential risks and mitigate them to the greatest extent possible, in order to more effectively achieve the goal of using digital imitation in the educational process.
2. In the case of such an experiment, inform students conscientiously about their interaction with AI, as well as explain the goals and objectives of using this tool. This will help students better understand how technology works and see its benefits for the educational process.

Practices:

In December 2023, a European Union **project on the use of digital twins in higher education institutions was launched**, which was joined by 11 universities in various countries[199]. The aim of the proposal is to expand the capabilities of higher education institutions in augmented reality (AR) and virtual reality (VR) through the use of digital twins. The project will prepare instructors who in the future will train teachers to work effectively and ethically with these tools.

Research on the issue:

1. University of Hong Kong researchers highlight qualities of the teaching staff that cannot be replaced by AI[200], including digital imitation. For example, educators bring real-world context by sharing examples and experiences that help students better understand learning material and connect it to life situations. They create space for discussion by presenting diverse perspectives, asking difficult questions and developing critical thinking, which AI is not yet able to fully achieve. In addition, teachers play an important role in conflict resolution, teaching peaceful settlement skills and promoting social responsibility, which makes them indispensable in educational practice.
2. NTT Technical Review researchers argue that **the use of digital twins by teachers should be transparent to students**. In order not to mislead students, they should know from the outset that they are communicating with AI[201].
3. A scientist from the Kyoto University of Foreign Languages pointed out **the advantages of using digital imitations**[202]. For example, this technology makes it possible to personalize content, taking into account the characteristics of the student audience (such as deaf-mute students or foreign students who do not speak the language well).
4. A group of researchers from Fujian Medical University noted that **the use of digital copies of teachers allows students in remote areas to access educational content** of the same quality as those in locations where many educational resources are available[203].

What do the experts think?

Sergey Roshchin,
Vice-Rector for Academic Affairs of the Higher School of Economics

"This technology may well be used for educational purposes. A lesson with a digital imitation of a teacher is like a lesson with a video recorded by the educator. It's just an avatar. However, if this format is chosen for a class, then students should be made aware that they are looking at an imitation of the teacher."

Vadim Perov,
Head of the Ethics Department of St. Petersburg State University

"Education is a mutual process. Therefore, firstly, from an ethical point of view, it is not enough to inform students about the "digital imitation" of a teacher, but it is necessary to obtain their consent. Secondly, if we recognize that a "digital teacher" is ethical, then the question arises about the ethics of the presence of "digital twins" of students in the audience[204]."

26 Is it ethical to use AI to write course assignments or other academic papers?

Answer:

Yes, if it does not contradict the principle of academic integrity, as well as local regulations of an educational or scientific organization. It is ethical to use AI technologies as a tool that allows you to process information and edit the texts of scientific papers, but not to substitute authorship. The responsibility for the accuracy of the data used and the final result is always borne by the person.

Justification:

- In its report 'The Era of AI in Higher Education', UNESCO highlights **the ability of AI to browse a large amount of literature** in order to quickly find the most relevant and up to date research[205]. Such systems use information from the internet, which may be unreliable and require rechecking.
- UAE University scientists believe that **AI can become a useful tool for teachers when marking student papers**[206]. It can make it possible to get an alternative opinion and quickly determine in which areas students need additional attention.
- According to **the UNESCO Guidelines on the Use of GenAI in Education and Scientific Research**, AI is best used for automated information collection and the preparation of a structure for scientific research. Possible risks outlined include the possibility of creating false information, for example, the use of non-existent research publications[207].
- **AI can be useful for devising works** according to a standard established by the verifying party and when checking the text for plagiarism, errors and linguistic stylistic inconsistencies.

Recommendations:

For an educational organization:

1. **Develop clear rules for students and teachers on the use of AI in the learning process**. This will help to avoid possible problems with plagiarism and ensure compliance with ethical standards.

2. **Encourage students to analyze information on their own and formulate their own conclusions.** This will contribute to the development of critical thinking and data skills.

3. **Provide students with access to high-quality sources of information and resources.** This will help them to conduct research and write scientific papers independently.

For students:

1. **Follow the rules set by the educational organization.** In some organizations, local regulations prohibit the use of AI in the preparation of works.

2. **Critically analyze and verify the factual basis of the answers offered by a neural network.** Neural networks can hallucinate and make mistakes, so additional verification will allow you to identify and eliminate possible inaccuracies.

3. **Use materials created with the help of AI only to reinforce your own scientific position.** They should not replace the main arguments.

Practices:

1. In 2023, a student in Moscow successfully defended a thesis written in 23 hours using a neural network. The student used ChatGPT to make a plan for the work and write the introduction and section on theory. However, the student spent 8 hours editing the text and writing the practical part of the thesis [208].
Neural networks can automate the process of searching for sources and information or check texts for spelling errors. However, it is worth remembering that AI is not a complete substitute for the thought process. Neural networks can create a draft of a scientific paper, but the creative part will still have to be performed by person.

2. **Many universities already adopt provisions on the use of AI in writing scientific papers.** For example, in May 2024, the Higher School of Economics (HSE) adopted 'Regulations for checking written academic papers for plagiarism and the use of generative models' [209]. According to Section 3 of these Regulations, failing to mention of the use of generative models is considered as a violation of academic rules.
The Moscow City Pedagogical University (MGPU) follows a different approach. In August 2023, at a meeting of the MGPU Academic Council, it was decided to legalize the use of AI technologies by students for the preparation of final qualifying papers. It means that students can use chatbots and other AI tools to obtain data and texts while working on graduation thesis [210].

What do the experts think?

Elena Bryzgalina,
Head of the Education Philosophy department at the Lomonosov Moscow State University (MSU) Faculty of Philosophy Moscow State University, Head of the Bioethics Master's Program

"Ethical aspects of behavior in the scientific and educational space include issues of compliance with author's ethics. Using feedback received from an AI tool has been called "AI plagiarism". Providing texts of educational and scientific works generated by AI tools under your own authorship, without indicating the use of such tools, can be qualified as academic fraud. The author of the work should be held responsible for violating author's ethics in educational and scientific situations."

Ivan Karlov,
Head of the Laboratory of Digital Transformation of Education at the HSE Institute of Education

"It is important to understand the purposes for which artificial intelligence is used. You can use it in different ways. You can ask it to write a term paper, or you can, as even experts are doing now, give an AI thesis and ask to give it in literary or scientific language. That is, when you actually already have all the work, and you use this tool to prepare the text of this work. In any work, there should be a part of the research that implies that a person does something with their own hands, and in any case, artificial intelligence will not do it for them[211]."

27 Is it ethical to use AI to check the work of students?

Answer:

Yes, AI can be used effectively as an auxiliary tool for teachers when checking student papers, automating routine tasks. However, the final decision should always be made by the teacher.

Justification:

- UNESCO, in its report on the use of **AI in education, notes that AI eases the burden on teachers**. If using it to perform routine assessment tasks, then the teacher will have more time to check the creative part of the tasks, as well as to give personalized feedback [212].
- The University of Oxford allows its students to **use AI self-testing tools to make better use of their study time**. AI is able to take into account the context and criteria for evaluating texts, work with different data formats, and provide personal recommendations [214].
- According to a study published in Data Science Central, **automatization of verification allows you to get results much quicker**. The teacher, in turn, can discuss them with the student later in real time [216].
- The Princeton Review, an international company providing services for entrance exam preparation, claims that **the verification of standardized tasks using AI allows them to be evaluated objectively, without bias or the 'personality factor' in the assessment process**, following pre-defined algorithms and criteria [213].
- A group of researchers from India believes that **AI mechanisms embedded in anti-plagiarism systems contribute to the observance of the principle of academic integrity**. They can analyze linguistic patterns, syntax, and semantic structures to identify cases where students have tried to hide plagiarism by changing the wording and structure of the source text [215].
- According to scientists from the University of London, **AI makes it possible to identify gaps in a student's knowledge**. For example, in addition to determining whether a student gave the correct answer or not, the AI can analyze the work to help teachers understand the thought process behind the student's answer [217].

Recommendations for teachers:

1. **Ensure the confidentiality of student data.** It is necessary to ensure the security and protection of a student's personal information, as well as to comply with personal data protection legislation and ethical principles.
2. **Inform students about the participation of AI in the verification of students' work.** This will increase their trust in both AI systems and the educational organization itself.
3. **Observe the principles of academic honesty, openness and respect for the individual.** For example, it is important to explain the assessment criteria to students so that they understand what the results are based on.
4. **Keep in mind that AI is your tool and assistant, not an expert.** The teacher is solely responsible for the assessment results.
5. Use national linguistic models so that there are no mistakes.

Practices:

1. **In China, teachers are already actively checking student test papers using artificial intelligence algorithms**[218]. The ZipGrande neural network, the key task of which is to quickly check the work of schoolchildren, already has 800,000 users.
   The program works as follows. The user points a smartphone camera at paper records, after which the AI checks the work for errors in just a few seconds and provides the result.
   As the survey showed, 60% of teachers believe that marking tests is the most time-consuming task, and therefore this system greatly simplified their work.

2. **The Government of the Russian Federation also intends to involve AI systems in checking homework in schools and for planning educational programs by 2030**[219].
   In June 2024, Denis Gribov, Deputy Minister of Education of the Russian Federation, speaking at the 2nd 'Shaping the Future' International Forum for Ministers of Education, said that special digital assistants were already being created for this purpose. Gribov noted that this project will also solve the problem of reducing the bureaucratic burden on teachers[220].

What do the experts think?

Sergey Roshchin,
Vice-Rector for Academic Affairs of the Higher School of Economics

“Of course it can. But before that, we must make sure that the AI makes mistakes no more often than a ‘real’ teacher. It’s like a simulator, only designed for the evaluation of the result.”

Dmitry Zubtsov,
Head of the Academy of Technology, Data and Cybersecurity, Sberbank University

“If the work shows the student’s knowledge of a particular issue, for example, their knowledge of the language, or understanding of certain terms, then this can be quite simply transferred to AI, perhaps in the mode of a decision-making assistance system for the teacher (highlighting incorrect answers). If the work is creative or contains analysis and conclusions that AI cannot always cope with, then the number of errors will be too significant and it is wrong to assign such a task to AI.”

Sergey Valyugin,
literature teacher of the ‘Nika’ School, lecturer at the Department of World Literature of the Pushkin Institute of the Russian Language, winner of the “Teacher of the Year at Moscow-2023”

“It is especially important to use AI when checking written works for compliance with spelling and punctuation norms (dictation, presentation, essays). But it is important to remember that in the presence of morphological homonyms (distinguishing conjunctions and introductory words, adverbs and nouns) AI does not always correctly take into account the context and additional verification by the teacher is required.”

28 Is it ethical to use AI-based proctoring systems?

Answer:

The use of proctoring systems on an ongoing basis is seen as unethical and impractical. At the same time, targeted application of proctoring to conduct key control measures may be ethically acceptable, provided that it is clearly regulated and the principle of non-discrimination is respected.

Justification:

Proctoring is a procedure for monitoring the progress of a remote test [221].

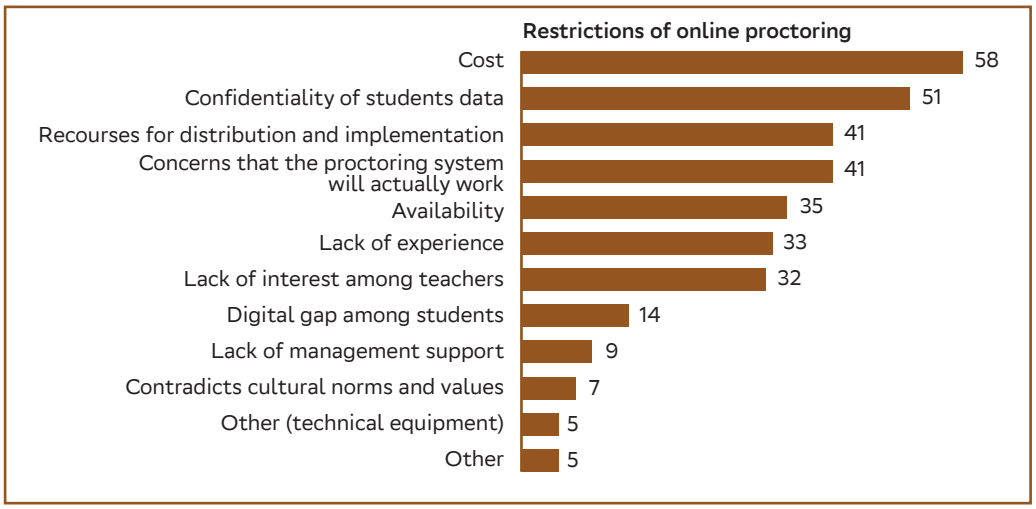
- **Constant proctoring carries risks of violation of fundamental rights.** Constant proctoring can be a serious interference in the personal lives of students, turning an educational environment into a 'panopticon' [222].
- According to a study published in the Journal of Information Technology, **the use of proctoring can lead to increased social inequality.** For example, in the case of people on low incomes who may not be able to afford suitable technical equipment [223].
- Scientists from the University of Melbourne believe **that constant proctoring contributes to a growing distrust of important social institutions.** Total control undermines trust between students and teachers, disrupting the psychologically comfortable atmosphere necessary for effective learning [224].
- **The limited use of** proctoring for key control activities is justified by the need to ensure equal conditions for all students and the objectivity of assessment. However, according to UNESCO, such decisions should be transparent to students and subject to appeal in controversial cases [225].
- **Strict control can demotivate** students, is detrimental to the development of independence, responsibility and conscientiousness as sustainable personal qualities.
- The OECD, in its report 'Online Exams in Higher Education during COVID-19' [226], highlights **the disadvantages of online proctoring**, for example, it increases student anxiety when taking exams, since they fear punishment due to technical failures.

Recommendations for educational organizations:

1. **It is recommended to seek a reasonable balance between the need to ensure academic integrity and the imperatives of respecting student autonomy and privacy.** Proctoring algorithms should be open to audit, set up to minimize errors and bias, and their decisions should be subject to appeal and human review in controversial cases.
2. **Excessive control through proctoring is not allowed.** Such an approach cannot replace efforts to create an atmosphere of trust and cultivate standards of integrity in an academic environment.
3. **Comply with ethical standards and legislation on the protection of personal data.** The use of proctoring systems requires clear regulation, guarantees of data confidentiality and protection against discrimination in accordance with the norms of information ethics.
4. **Prioritize educating students about conscious academic integrity.** Ethical codes, trainings and engaging forms of education can help in achieving this goal.

Research on the issue:

In April 2020, not-for-profit organization Educause conducted a study to identify the main challenges of distance education and potential ways to solve emerging problems.
According to a survey, the main problems faced by educational institutions during the implementation of proctoring systems are **the cost of online monitoring** (58% of respondents), as well as **ensuring the confidentiality of student data** (51%).



Source: NPO Educase

What do the experts think?

Farida Mailenova,
Leading Research Fellow, Institute of Philosophy, Russian Academy of Sciences

"Proctoring in the performance of control tasks, online tests is necessary, as it increases the degree of fairness and objectivity in the assessment of results; and in general contributes to increasing responsibility among students' [227]. An important ethical point is that students should be aware of this. There is no special need to use it on an ongoing basis, since the specifics of online learning makes it possible to view recordings of lectures, while the students are themselves responsible for the learning outcome."

Dmitry Istomin,
CEO at Examus

"Any system is imperfect. The drawback of proctoring is precisely that it needed to be invented. For some reason, people don't approach exams honestly.
The proctoring system brings to education, equality, first and foremost, something that is so often talked about. You can continuously improve the system and the accuracy of the algorithms. If you look at other industries where automation and recognition are used, like in cars and drones, then you can see this is an endless process of improvement' [228]."

29 Is it acceptable to lower a student's grade if it is suspected they have used AI in their work?

Answer:

Lowering ratings just because of the detection of signs of AI use seems too categorical, but ignoring the facts of AI use is also wrong.

Justification:

- The Massachusetts University of Technology notes that **sanctions that do not take into account the actual contribution made by the student** can lead to fear and suppression of creative initiative among students' [229].
- **According to a Stanford University study, false positives are most often associated with lexical features of the text' [230].** Such systems often identify non-native texts as being generated content, as native speakers usually have a larger vocabulary and a better understanding of grammar. Non-native speakers write using the most common phrases. The same is true of generative AI. In fact, it simulates human writing based on all of the data it has ever processed.
- **Lowering a grade for using AI without clear evidence of dishonesty** undermines the relationship of trust between student and teacher. In turn, the effectiveness of learning directly depends on this factor.
- **Algorithms for detecting the use of AI can give false positives.** It is not necessary to rely solely on the results of these systems, as this could lead to the punishment of the innocent and the rewarding of the guilty.
- **The teachers' focus on identifying AI distracts from meaningful feedback and discussion of the essence of the work.** This can negatively affect the development of the discussion and argumentation skills of students.

Recommendations for educational organizations:

1. **Work out a differentiated approach to the possible use of AI by students**. Such an approach should be based on an open dialogue with all stakeholders.

2. **Formulate and consolidate clear and transparent rules for the use of AI.** Reflect within these rules the criteria for evaluating works created using generative AI tools.
3. **Take preventive measures.** For example, explain to students the ethical principles of working with AI. Teach them about responsible interaction with this technology, cultivating the values of academic integrity.
4. **Lowering grades is possible as a last resort in cases of proven abuse.** However, lowering grades should not be an automatic reaction in cases where AI was used.
5. **The evaluation system should be set up to encourage the ethical use of AI to solve tasks.** Adaptation will require experimentation, constant feedback, and the willingness to flexibly adjust approaches.

Research on the issue:

In July 2023, Stanford University conducted a study: scientists evaluated several publicly available systems that claimed to be able to recognize generated text, using samples written by native and non-native English speakers.
As a result, 89 out of 91 (97.8%) essays written by non-native speakers were marked as AI-generated by at least one of seven different tools.
To test the hypothesis that limited vocabulary contributes to bias, scientists used ChatGPT to 'enrich' the language, seeking to mimic the use of native speakers' vocabulary.
This intervention led to a significant reduction in the previous, erroneous classifications. The average level of false positive results decreased by 49.45% (from 61.22% to 11.77%) [231].

Practices:

OpenAI shut down its own AI detector in July 2023 after discovering it had a "low level of accuracy." A post on the company's website reported that **'none' of the generated content recognition systems, including their own, "have proven that they can reliably distinguish AI-generated content from human-created content."**
OpenAI noted that the existing systems have a clear bias against students who study English as a second language, as well as students whose texts are particularly formulaic or concise.
Moreover, according to the company it is very easy to bypass AI recognition systems by simply adding a few common sentences [232].

What do the experts think?

Yury Chekhovich,
Executive Director at Antiplagiat

"When the Anti-Plagiarism system detects that there are many signs in the text that it was written by a neural network, it highlights a certain fragment of text as suspicious. However, it is not yet possible to draw a final conclusion that this text was written by a neural network. Our system acts only as a tool that highlights suspicious fragments of work, and then it's up to the person to decide."

Olga Frantsuzova,
Department of Philosophy of Education,
Moscow State University

"Recently, teachers have been faced with situations demonstrating a lack of independence in the work of pupils and students. The weakness of the tools for detecting deviations from the rules and the wide availability of technologies has cultivated poor quality of outcomes. Of course, learners should not be punished for the "mechanization" of skills in searching for literature as well as design, raising issues and moving forward tasks or hypotheses. But one should be more careful with the work content, ensuring violators are held accountable, while at the same time educating and familiarizing them with the academic ethics of creating their works."

Alexander Gasnikov,
Rector of Innopolis University

"The question of using AI in academic works is not as simple as it might seem at first glance. The initial reaction is that it is necessary to ban all this and not allow it to be used, otherwise people will not learn anything... However, on the other hand the ability to properly use AI to solve a particular task, including at the learning stage, can in turn be an element of learning and useful in the future. The solution may be the division of tasks into those in which it is allowed (and even recommended) to use all available means, including those based on AI, and those tasks in which it is prohibited... Violations in these cases can probably be identified as cheating."

30 Is it ethical to limit the use of AI by children for educational purposes when they are outside the relevant educational institutions?

Answer:

Do not restrict the use of AI by children for educational purposes outside of the relevant institutions completely. However, such use should take place under adult supervision, taking into account age restrictions and the level of development of children.

Justification:

- The UK Department of Education believes that **AI can be a useful tool for children's learning and development**. Students can familiarize themselves with educational materials outside the classroom, and then come to class with basic knowledge to participate in more interactive activities [233].
- UNICEF, in its report 'Policy guidance on AI for children', emphasizes that **AI technologies can be used as an assistant** in the process of completing homework, developing additional skills (for example, creative ones), including for children with disabilities [234].
- **AI can produce unwanted or unreliable content.** Reasonable restrictions on the use of AI by parents and filtering the content by the developers, reduces the risk of them absorbing this kind of information, protects them from negative impact and safeguards their mental health.
- According to a study published by High Speed Training, **specialized AI systems for children of different ages can help to understand disciplines and life topics that are not explained in educational institutions.** For example, programs can provide insight into psychological concepts and theories, helping children develop an understanding of human behavior and emotions [235].

Recommendations:

Recommendations for developers:

1. **When developing AI solutions for additional education of children, take into account age restrictions.** Developers should differentiate between content that will correspond to the level of development and mental resilience of children.
2. **Integrate similar functionality into your services.** Developers of services based on generative models should establish the possibility of parental control and age restrictions on viewing generated content.

Recommendations for parents:

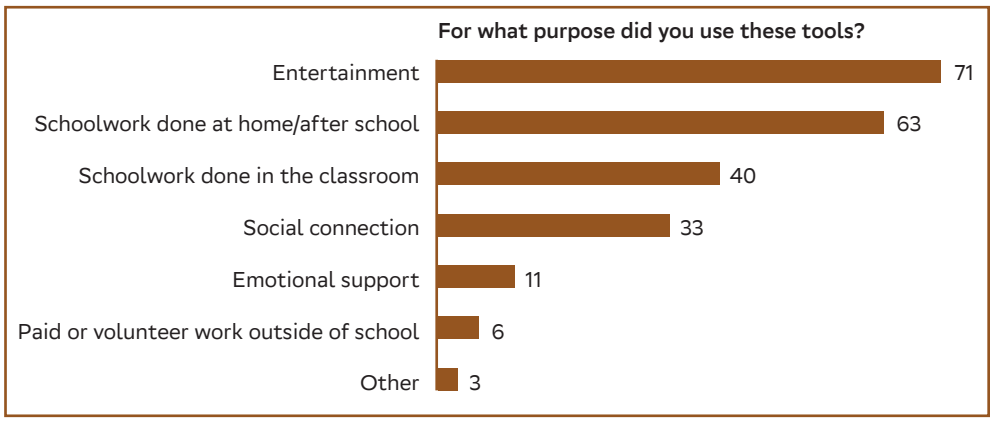
1. **Monitor the use of AI for educational purposes by school-age children.** Children may abuse technology in order to complete homework and not build up the necessary knowledge and independent skills.
2. **Be responsible about choosing AI-based services for your children.** Choose those developers who provide information about their algorithms and values as openly as possible, as well as those that specifically focus on children's education.

Research on the issue:

In February 2024, the results of a survey conducted by Hart Research were published. The survey was devoted to the use of artificial intelligence among teenagers.

58% of respondents said that AI helps them improve their academic performance at school, and also promotes interest in additional learning outside educational organizations.

Young people are particularly inclined to use generative AI: 60% of respondents admitted that they use genAI tools on a regular basis.



Source: Hart Research [236]

At the same time, 63% of the surveyed children aged 9 to 17 years use these tools specifically for educational purposes, including for homework [236].

What do the experts think?

Ilya Pomerantsev,
Head of the AI department at Globus IT

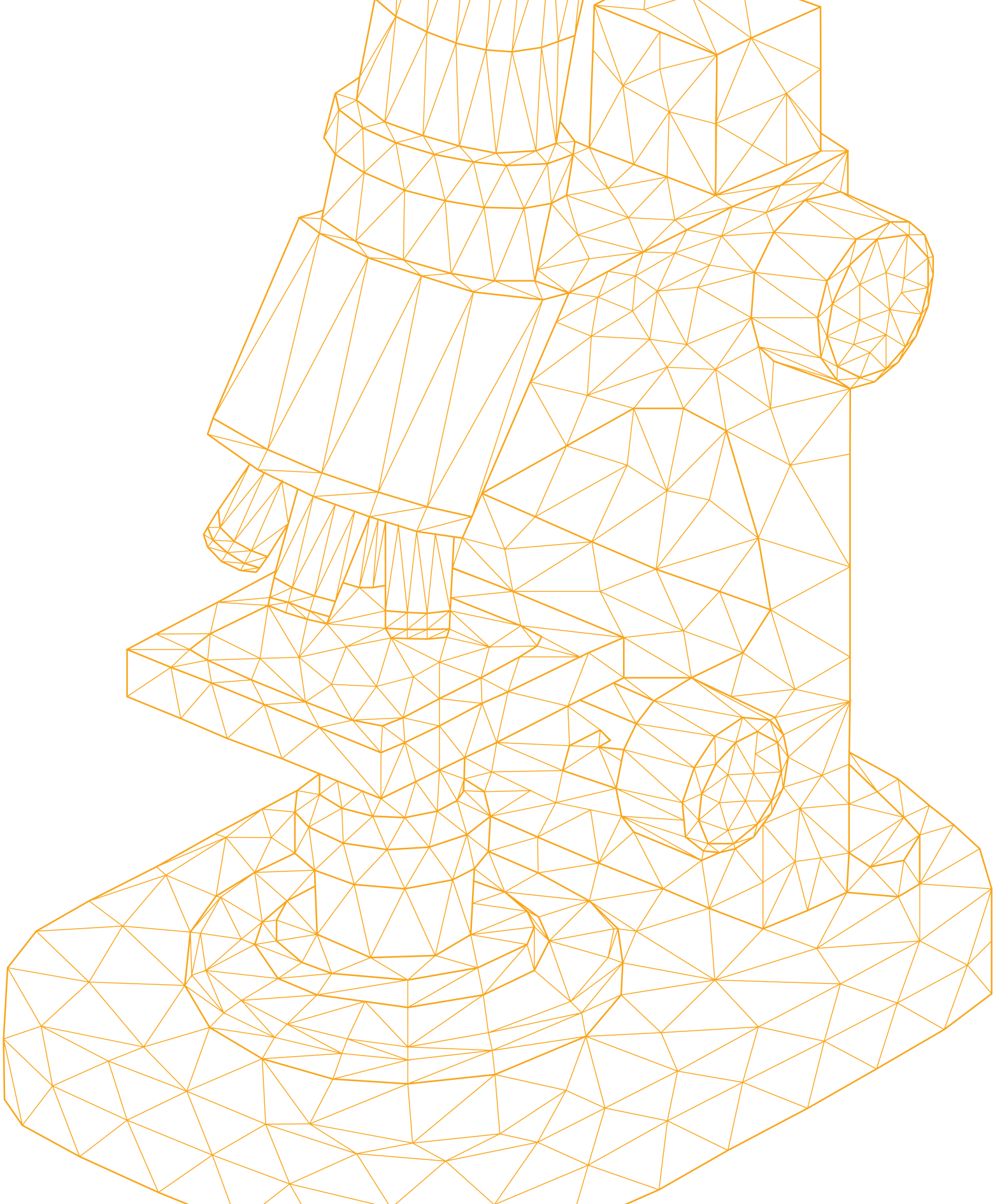
“The inclusion of AI technologies in the learning process implies a gradual transformation of the education system. This also applies to approaches to the presentation of information, its assimilation and verification. It is important to take into account age restrictions and use specialized solutions, including parental control tools. It is undesirable for children to use publicly available, non-specialized AI-based solutions for educational purposes outside of educational institutions.”

Alexey Khabibullin,
Head of the Directorate for Pre-University and Olympiad Training, Neymark IT Campus

“It is necessary to create and implement special programs for teaching teachers and students at pedagogical universities as well as parents about the opportunities that neural networks can provide in the education system and for development of a child. Limiting the use of AI technologies, if it’s necessary, should be under adult supervision.”

Chapter 06

AI and medicine



31 Is it ethical for a person to self-medicate with AI?

Answer:

To support health decisions, it is possible to use only specialized and certified AI systems that are designed for this and tested for accuracy. Ordinary chatbots that are not intended for medical purposes do not have the necessary reliability and may give incorrect recommendations, so their use is unsafe.

Justification:

- A group of American researchers believe that **AI technologies can partially solve the problem of a shortage of specialists and create additional opportunities for residents of hard-to-reach areas**[237]**.** In many rural areas in developing countries, there are few qualified doctors, and a large number of patients need the help of nurses or nursing staff.
- **The ability to use specialized AI systems to help with health issues allows patients to receive prompt recommendations** and information about their health condition.
- In Russia, a study was conducted when a chatbot without a medical specialization was asked the same question indicating different roles of a doctor. In one case, being assigned the 'role' of gastroenterologist, the bot offered diagnoses including acute appendicitis, pancreatitis and cholecystitis. However, when the bot took on the role of gynecologist, it offered other diagnoses, for example, PMS or ovarian cyst. This shows that **ordinary bots are keyword-oriented and are not able to assess the full clinical picture**, which makes them unreliable for medical use.

Recommendations for developers:

1. **Analyze potential risks when developing medical AI systems.** It is important to consider the possible consequences of providing false information.
2. **Develop systems based on ethical principles of safety and fairness.** In case of detection of life-threatening conditions, the system should recommend immediate medical attention, observing the principles of medical ethics and patient well-being.

Recommendations for users:

1. **Use only specialized AI systems to make decisions on health issues.** Such systems are designed taking into account medical standards and have high reliability, which reduces the risk of errors.
2. **Always consult with medical professionals if you are concerned about your health.** Even proven AI tools do not replace a doctor's professional opinion and should be used as additional support, not the main source of guidance.

Research on the issue:

In June 2024, KFF, the leading organization in the health policy in the United States, conducted a survey among American citizens on the use of chatbots to obtain medical information. According to the study, about **one in six adults (17%) says they use AI chatbots at least once a month to get medical information and advice**, and this figure reaches a quarter (25%) among adults under the age of 30. The majority of adults, including the majority (56%) of those who use or interact with AI, are unsure of the accuracy of the medical information provided by AI-based chatbots [238].

Practices:

In 2021, the Laboratory of Computer Science and Artificial Intelligence at the Massachusetts Institute of Technology developed an **AI tool to track the correctness of medication intake, as well as reminders and the forwarding of the data to the doctor.** A wireless sensor was installed at the patient's home. The AI system continuously and automatically analyzed radio signals and documented the results, which were uploaded over the internet and added to the patient's digital medical record. Reminders were sent to the patient if they did not take their medicine at the appointed time. Authorized medical professionals also had access to these records to track the condition of patients [239].

What do the experts think?

Diana Khasanova,
Associate Professor of the Department of Digital Technologies in Healthcare at Kazan State Medical University, CEO of Brainphone

"The availability of medical care varies in different regions of Russia, which affects the quality and life expectancy of the country's population.
AI tools can help align the possibilities of medical care, especially in hard-to-reach regions, which is among the healthcare priorities given the long distances and aging population in the country."

Pavel Vorobyov,
Professor, Chairman of the Board of the Moscow City Scientific Society of Therapists

"Residents of many thousands of villages in the country are deprived of contact with medical workers. There is nothing available there except for reception staff and medical commissioners who do not have a medical education. That's why they need decision support systems, including those based on AI, to help them build an adequate mechanism for both emergency and planned assistance and support."

32 Is it ethical for a doctor to delegate decision-making on prevention, diagnosis, treatment and rehabilitation to AI?

Answer:

As a general rule, no. The use of AI in diagnosis, treatment and rehabilitation can be considered ethical when its recommendations are checked and confirmed by qualified specialists. AI can help doctors by preparing conclusions and recommendations, but the final decision must be made by a human.

Justification:

- As noted in the strategic document of the European Parliament 'Robots in healthcare: a solution or a problem?' [240], there is currently insufficient experience, and **the existing regulatory and ethical frameworks do not fully eliminate risks to patient safety**, which is necessary to build trust and acceptance by users, both patients and non-patients.
- Researchers at Khalifa University of Science and Technology (UAE) believe that **the problem of the 'black box' — the growing ambiguity and complexity of the interpretation of algorithmic functions**, in terms of both the learning process and the reliability of the results — creates serious obstacles to delegating decision-making to AI systems. It also does not meet the ethical criterion of explainability — covering why a particular medical decision was made [241].
- **Patients may feel distrust and discomfort knowing that their treatment is completely controlled by AI.** AIS monitored by a doctor provides a higher level of support and trust and reduces risks.
- A group of scientists from Switzerland and the United States recall that **AI can demonstrate potential bias against certain groups of patients,** for example, due to insufficient training data. This can lead to discrimination and the violation of the bioethics principles [242].
- **The complete transfer of decision-making to AI can lead to a decrease in human control and the sense of responsibility** — risking of loss of autonomy in decision-making.

Recommendations for developers:

1. **Take into account the principles of data security, reliability and confidentiality.** This minimizes the risk of discrimination and abuse.
2. **Develop AI systems with the ability to explain and make decisions transparent**. This will allow medical professionals to better understand the logic of AI and increase confidence in its recommendations.

Recommendations for medical professionals:

1. **Use AI as a decision support tool, but make the final decision yourself.** It is important to remember that AI provides information and does not replace medical experience.
2. **Check and evaluate AI recommendations in the context of each specific case.** Adapt the AI's suggestions, taking into account the individual characteristics of the patient and the clinical situation, in order to avoid mistakes.

Research on the issue:

Pew Research Center in Washington D.C. has studied public opinion about AI in healthcare and medicine. Six out of ten American adults say they would feel uncomfortable if their doctor relied on AI to diagnose diseases and recommend treatment methods. The survey also showed that only 38% believe that the use of artificial intelligence to diagnose diseases and recommend treatment methods will lead to an improvement in the health of patients [243].

Practices:

In Russia, AI is already actively used in healthcare. In 2023, Russian regions purchased 106 medical devices enhanced with artificial intelligence with a total cost of around 448.43 million rubles. These technologies have been implemented in 85 regions of the Russian Federation [244].
Moreover, by the end of 2023, 22 million medical records in Russia have been analyzed using AI tools. Voice document-filling services are used in six regions, and AI virtual assistants are used in 29 regions for making appointments with doctors.

What do the experts think?

Bulat Magdiev,
Sechenov University Medical Device Research Center

"The use of AI in healthcare may be ethically acceptable, provided that AI acts as a physician's assistant and does not replace them completely. The combination of AI capabilities and a doctor's expert opinion can improve diagnostic accuracy, optimize treatment and improve patient outcomes, while maintaining human control and responsibility."

Elena Bryzgalina,
Head of the Education Philosophy department at the Lomonosov Moscow State University (MSU) Faculty of Philosophy, Head of the Bioethics Master's Program

"AI performs only the role of an assistant, as a medical decision support system. The use of AI systems in medicine could result in possible harm to patients. The definition of liability for harm caused relates to analysis of the actions of the person — a doctor or a medical institution using AI systems as tools to support medical decisions. Delegating decision-making and assigning responsibility to AI is impossible."

Boris Zingerman,
Director of the Association of Developers and Users of AI in Medicine "National Medical Knowledge Base"

"At present, there are very few autonomous artificial intelligence solutions registered around the world, there are no more than 10 of them. Nevertheless, they do exist and are likely to be important in the future. That is, they should of course be checked with much greater reliability than those solutions where, after all, a person reflects the final result, but these are the areas that are fundamentally important to us."

33 Is it ethical to use AI to handle telling a patient bad news?

Answer:

No, using an AI system to convey 'bad news' can be considered unethical, since this method of communicating medical information could be traumatic for the patient.

Justification:

- Receiving a diagnosis of a serious illness causes strong emotional reactions such as shock, fear, anger or sadness. **An actual doctor can assess the patient's condition, offer the necessary support** and, if necessary, invite the relevant specialists.
- According to a study published in the Health Care Science journal, **reporting bad news requires tremendous skill and caution**, as patients often experience symptoms of anxiety and depression after they have been given with a frightening diagnosis, and various recommendations for reporting bad news have been developed to minimize psychological harm to patients. Recommendations that a doctor should follow when reporting 'bad news' include protocols like SPIKES, BREAK and FINE, among others. **In theory, an AIS can be taught these principles too**[245].
- **It is important not only to convey the information, but also to make sure that it is understood correctly.** A doctor is able to answer questions, clarify details and support the patient, which the AI may not be able to do properly.
- Indian doctor Ligi Thomas believes that in the case of the widespread introduction of AIS and AI-based chatbots, **doctors may lose their communication skills with patients in difficult situations and then avoid such communication**. Patients who are made to feel alienated by their attending physician would start resorting to self-diagnosis and self-medication[246].

Recommendations for using AI to report ‘bad news’:

1. Doctors are not recommended use AI chatbots to replace live communication with patients. However,this rule has exceptions, for example, AI can be useful for supporting patients after a doctor reports ‘bad news’, for example, it can remind them about taking medications or accompany the patient during the rehabilitation process.
2. Developers should consider the possible emotional and psychological consequences for patients and doctors. The introduction of automated systems for communication with patients should be based on a thorough analysis of the benefits and risks.
3. Implement principles of medical ethics in the use of bots to inform patients. These principles should take into account the needs of the patient and provide the opportunity to interact with a doctor.

Research on the issue:

In 2023, American doctors decided to conduct an experiment and asked ChatGPT to help them communicate with patients more sympathetically. According to the results of the study, it turned out that **the answers created by the program turned out to be more empathetic than those from real doctors**.
Based on the data, studies were conducted where medical experts were asked to compare how doctors and ChatGPT conveyed bad news to patients. It turned out that 78.6% of the people surveyed preferred the answer generated by AI [247].

Practices:

In 2019, **a doctor at a California clinic entrusted a robot to inform a patient about a serious diagnosis.** The patient was unprepared for the information to be conveyed in such a way and was shocked, as were his relatives. Wilharm, the patient’s granddaughter, told reporters: “I think they should have had more dignity and treated my grandfather better than they did.”
Her grandfather, 78-year-old Ernest Quintana, died the day after the diagnosis was announced [248].

What do the experts think?

Elena Grebenshchikova,
Director of the Institute of Humanities, Head of the Department of Bioethics of the Pirogov Russian National Research Medical University

"Conveying 'bad news' to a patient and/or their family members requires a special approach, sensitivity, consideration for emotional state, and a willingness to support a person in a difficult moment and demonstrate that this information is not the end of the line, after which they can expect doctors to turn their backs on the patient. A robot will not be able to do this fully, and patients will accuse the healthcare system of callous and inhumane treatment. In addition, the robot is unable to sense unwillingness or a lack of preparation of a patient to receive information, it will not be able to understand the context of the situation, which allows the doctor to choose the right words, the correct time and place for them to be able to handle to information. 'Bad news' about a child's health has a negative impact on the entire family. Parents will have questions, for example, related to other children in the family – whether they need to be informed about a situation or whether there is any threat to their health. Furthermore, a robot would not be able to understand that it is necessary to repeat the information, to make sure it has been adequately understood, by responding both to psychological aspects and to practical requests by parents."

Anastasia Ugleva,
Professor at the School of Philosophy and Cultural Studies, Deputy Director of the Center for Transfer and Management of Socio-Economic Information at the Higher School of Economics

"In itself, the message with 'bad news' is no different from the information contained in, for example, an electronic medical record on the results of an analysis or description of an appointment with the doctor. At the same time, a patient should have the right to choose with whom to communicate about the state of their health — whether a doctor, a conversational assistant or a chatbot. However, in my opinion ethical issues arise not so much at the moment that a life-changing diagnosis is pronounced with the right words of support (AI can cope with this well), but in connection with the need to control the medical and social consequences. If AI technology is able to assess in real time the risks of a sharp deterioration in the patient's well-being and/or occurrences like suicidal thoughts as a result of receiving troubling information, and then also be able to promptly provide immediate psychological support, then this use of AI should be recognized as 'ethical'. In other cases, the use of AI should be abandoned."

34 Is a separate consent needed from a patient for the use of AI in treatment?

Answer:

It seems ethical to disclose information to the patient about the use of AI by a doctor and obtain their consent to the use it within the general consent to carry out medical manipulations. At the same time, treatment should always be carried out by a person, and AI should act solely as a tool.

Justification:

- In Russia, **a patient must give informed consent before a medical intervention is performed** [249]. To do this, the medical professional must provide comprehensive information about the goals, methods, risks, intervention options, consequences and the expected results in an accessible form.
- Spanish scientists note that the provisions of the EU Personal Data Law on automated decision-making apply only when the decision is "based solely" on AI, which means that **in situations where AI is used as a decision support tool, there is no legal obligation to inform patients about its use** [250].

Recommendations for medical professionals:

1. **It is recommended that the information on the use of AI in the provision of medical care should be disclosed** to ensure transparency, responsibility and respect for patient autonomy.
2. **The procedure of signing informed voluntary consent should be considered a way to fully inform the patient,** rather than a formal procedure.
3. **It is recommended to update knowledge about the key aspects of AI in medicine** as necessary to adequately inform the patient.
4. **Regular events should be held to raise public awareness of the possibilities, limitations and basic principles of AI in medicine**, as well as the risks associated with it.
5. **It is important to explain to patients about who is responsible for the provision of medical care with the use of AI** and explicitly underline their right to refuse medical intervention.

Recommendation for patients:

1. **Before signing informed consent, ask medical professionals direct questions about the stages, methods and risks of providing medical care**, including with the use of AI.

WHO's approach:

The ethical recommendations of the World Health Organization indicate that:

- AI technologies should not be used to experiment or manipulate people in the healthcare system without valid informed consent;
- The use of machine learning algorithms in diagnosis, prognosis and treatment plans must be included in the process of obtaining informed and valid consent;
- The provision of basic services should not be restricted or denied if a person does not give consent, and neither the Government nor individuals should offer additional incentives or inducements to those willing to give consent [251].

Research on the issue:

Researchers at Hanyang University Law School in South Korea conducted a survey of 1,000 respondents to assess the importance of patients being informed about the use of AI in diagnosis when deciding on treatment. The survey results showed that people attach more importance to information about the use of AI in diagnostics compared to consulting with a human specialist, for example, a radiologist. This indicates that comparing AI consultation and human consultation does not reflect the whole picture and does not justify the practice of doctors not to disclose information about the use of AI to support decision-making [252].

The survey participants perceived information about the use of AI as more important or equivalent to the lower limit for regularly disclosed information, which emphasizes the need to provide information about the use of AI in diagnostic procedures. This confirms that disclosure of information about the use of AI in diagnostics is an important aspect of physician-patient interaction, contributing to increased trust and understanding of the treatment decision-making process.

What do the experts think?

Elena Grebenshchikova,
Director of the Institute of Humanities, Head of the Department of Bioethics of the Pirogov Russian National Research Medical University

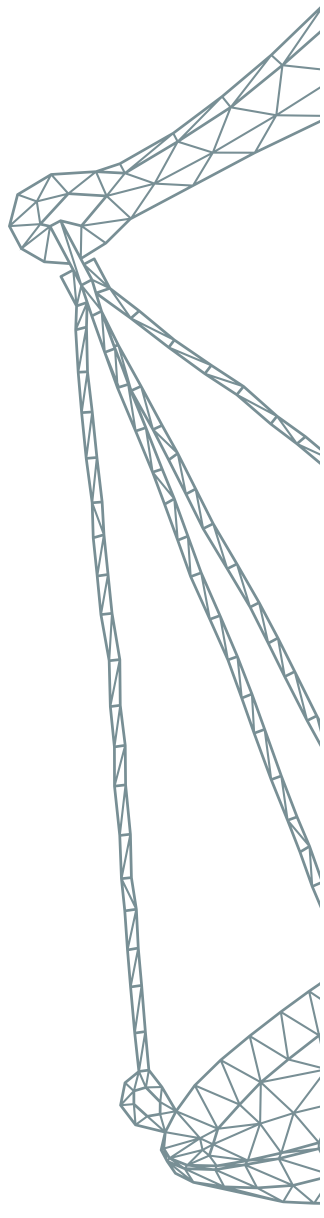
"The need for separate informed consent should be determined by its functions and the choice of the patient. For example, if AI is used by a doctor only for advisory purposes, then any decision by the doctor is only their choice, and a separate IDS is not required accordingly. But if, for example, the doctor suggested using AI for diagnostic purposes during a consultation, then the patient must be fully informed and sign an IDS form. The goal of implementing AI in healthcare is to improve the quality of medical care and to help both patients and doctors, which is impossible without taking into account the established norms of medical ethics, among which informed voluntary consent plays a key role."

Pavel Vorobyov,
Professor, Chairman of the Board of the Moscow City Scientific Society of Therapists

"The principle of reasonable sufficiency should be used. Bringing up for discussion with patient all the subtleties of the medical technologies used, including those using AIS would be completely redundant, since decisions in the existing health care system are made by a medical professional, and AIS plays only an auxiliary, albeit important role."

Chapter 07

AI in the judiciary



35 Is it ethical for judges to use AI?

Answer:

Yes, if permitted by law, AI is ethically and expediently used as a tool to increase efficiency, for example, when compiling a summary of a case or to automate and simplify the search for judicial practice.

Justification:

- Researchers at the University of New Hampshire believe that **AI can only act as an assistant judge**. The summary does not prejudge the assessment of evidence given by a judge in the process of establishing the circumstances of a case and applying the law, but rather assists the judge to quickly assimilate the evidence and exclude errors [253].
- **The synopsis will summarize the evidence, filter it according to certain criteria** — admissibility from a legal perspective, for example, indicating the absence of a notarized certificate where required, as well as relevance, for example, indicate that evidence is clearly irrelevant.
- According to researchers at Woksen University in India , **AI can help courts reduce the time taken to consider cases** by providing accurate information and analysis based on precedents. This speeds up the decision-making process and increases the accuracy and thoroughness of the legal assessment [254].
- A group of Indian scientists claims that **AI will expand the judge's ability to analyze judicial practice**: It will quickly select examples of relevant court decisions and a brief summary of legal positions. Natural language processing, machine learning, and data analytics have become indispensable tools for quickly reviewing legal documents, identifying necessary information, and predicting case outcomes [255].

Recommendations for judges:

1. The priority in making a decision on the possibility of using or not using AI, first of all, is the legislation and the positions of the highest judicial authorities on this issue.

2. Use AI to analyze voluminous data and search for use cases, but leave the final decision to yourself. The use of AI helps to process information quickly, but final judgment, as a rule, requires a human assessment of all the circumstances of the case.
3. The use of tools for generalization and summarizing case materials and judicial practice should be performed under the control of a judge. The judge should be able to check and correct the conclusions of the AI in order to avoid mistakes and ensure a fair hearing of the case.
4. At the stages of implementation and trial operation, the AI should be rechecked for the reliability of the information provided and the actual availability of relevant judicial acts and laws.

Practices:

1. In October 2021, **the French Court of Cassation launched the Judilibre** digital database containing 480,000 judgments rendered since 1947. Originally intended for judges and lawyers, Judilibre will gradually become available to applicants by 2025. The tool uses AI to optimize research and systematize court decisions. AI also makes it possible to pseudonymize data[256].

2. **In the USA, a model based on AI technology called Caselaw Access is used.**
This system includes a dataset of more than 6.7 million cases and makes it possible to determine the outcome of a case based on relevant precedents, judicial decisions and background statements from more than 400 courts.
Caselaw Access allows judges to quickly find cases relevant to that under consideration and take them into account when making a decision[257].

3. In May 2024, the Information and Communication Media Development Authority of Singapore (IMDA) announced a collaboration with the Singapore Academy of Law (SAL) to jointly develop a new large language model that will make legal research faster and more efficient. Known as the GPT-Legal model, it will be deployed on LawNet in stages from September 2024. In the first phase of implementation, **GPT-Legal will be used to summarize more than 15,000 Singapore court decisions**, providing brief information on keywords, facts and conclusions from judgments[258].

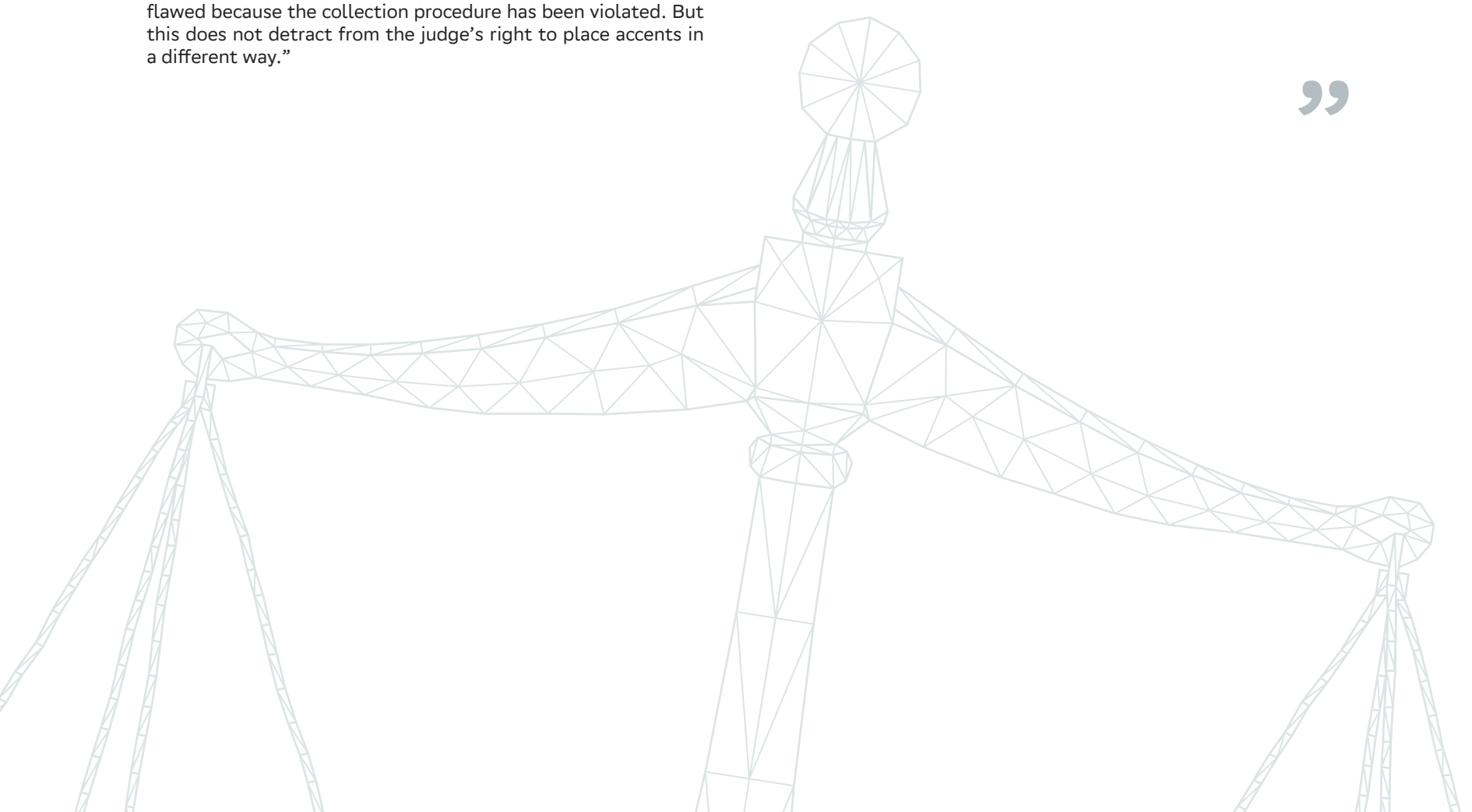
What do the experts think?

Elena Avakian,
Vice President of the Federal Chamber of Lawyers of the Russian Federation

"The summary of the case prepared by AI is important for the judge, because they may not study the entire volume of materials, of which not everything will have evidentiary value; and it is important for the parties that they understand what aspects the court has paid attention to and what they need to strengthen in their position. AI will be able to highlight the main problems of the collected evidence, for example, to indicate that proof is flawed because the collection procedure has been violated. But this does not detract from the judge's right to place accents in a different way."

Igor Yemshanov,
Chairman of the Commission of the Council of Judges of the Amur Region on automation and informatization of courts

"Given the time constraints placed on judges, the ability of AI to review texts would be very useful. The selection and intellectual retelling of decisions previously made by other courts would allow the judge to quickly get immersed into the subject and prepare for the case."



36 Is it ethical for parties in a court case to use AI?

Answer:

Yes, if the recommendations below are followed, AI can help overcome the problem of the lack of legal training of the parties: for example, describe the rules of jurisdiction and the procedure for going to court, help in drafting procedural documents or answer complex legal questions of citizens in simple language. At the same time, it is important to remember that, for such an application of AI, it is always necessary to take into account the requirements of legislation and the positions of judicial authorities on this issue (if any).

Justification:

- According to researchers from Suffolk University Law School, **AI can enhance access to justice by translating complex legal rules into layman's terms** and answering specific legal questions [259].
- **Using AI to legally substantiate an application produces a risk of factual errors that the system may make.** For example, guidelines prepared by Queensland courts in Australia on the use of Generative AI by non-lawyers make the applicant responsible for the accuracy and reliability of information received from the chatbot and submitted to the court [260].
- Researchers from Concordia University and the University of Montreal note that **access to justice is hampered by the high cost of consultations**. The work of AI chatbots will make it possible to save on paid legal assistance [261].

Recommendations for courts:

1. Inform users about the opportunities benefits and risks of using AI systems. By helping to overcome the formal procedures of going to court, these systems will increase the level of access that the general public has to the justice system.
2. Choose reliable and proven AI systems that comply with legal requirements and ensure data security.

Recommendations for developers:

1. **Follow the principles of transparency when developing AI systems.** Systems should be explicable and understandable to end users, especially when it comes to legal advice and answers to questions.
2. **Provide technical support and training for users.** Create detailed manuals so that citizens and employees in the judicial system can easily master working with AI systems and avoid common mistakes.

Recommendations for users:

1. **Check the legal arguments and recommendations proposed by the AI before using them.** AI sometimes makes mistakes that can be misleading. Always check the data with reliable sources and consult with lawyers if in doubt.
2. **Use AI as an aid tool for preparation.** Final conclusions and decisions should be based on consultations with professional lawyers.

Practices:

1. **In 2017, a robot named "Xiaofa" was put into operation at the Beijing People's Court.** The 1.46 meter tall robot provides consultations to visitors, answering complex legal questions in simple language. It can move its head and wave its arms when instructions appear on the screen, and direct people to the right window to receive court services.
   The AI-based tool is capable of answering more than 40,000 procedural and 30,000 legal questions. As a result of its implementation, it was possible to significantly speed up the process of applying to the court[262].

2. **In Arizona (USA), chatbots are actively used to automate justice.**
   So, a bot is used, which, at the request of the user, evaluates the likelihood of overturning a criminal charge. If the response is positive, the bot helps with filling out the petition and submitting it to the court[263].
   Another chatbot is specially designed to help with disputes arising from lease agreements. The bot is able to give step-by-step instructions on resolving rental disputes and provide recommendations on filling out procedural documents.

3. The first well-known **case with liability for presenting a legal position to the court with factual errors produced by AI** occurred in New York (USA). Preparing for a lawsuit, lawyer Stephen Schwartz used ChatGPT to search for precedent cases. During the trial, it turned out that the chatbot had fabricated these cases and even indicated that the non-existent decisions were made by the current judges. Now the lawyer has to pay $5,000 and notify each of the judges whose names appeared in the fictitious materials[264]. Unfortunately, this is not the only case in the United States when lawyers did not check the reliability of answers by the neural network. Therefore, in July 2024, **the American Bar Association issued ethical guidelines** on the use of GenAI in professional activities[265].

What do the experts think?

Vladimir Yarkov,
Head of the Department of Civil Procedure of the Yakovlev Ural State Law University

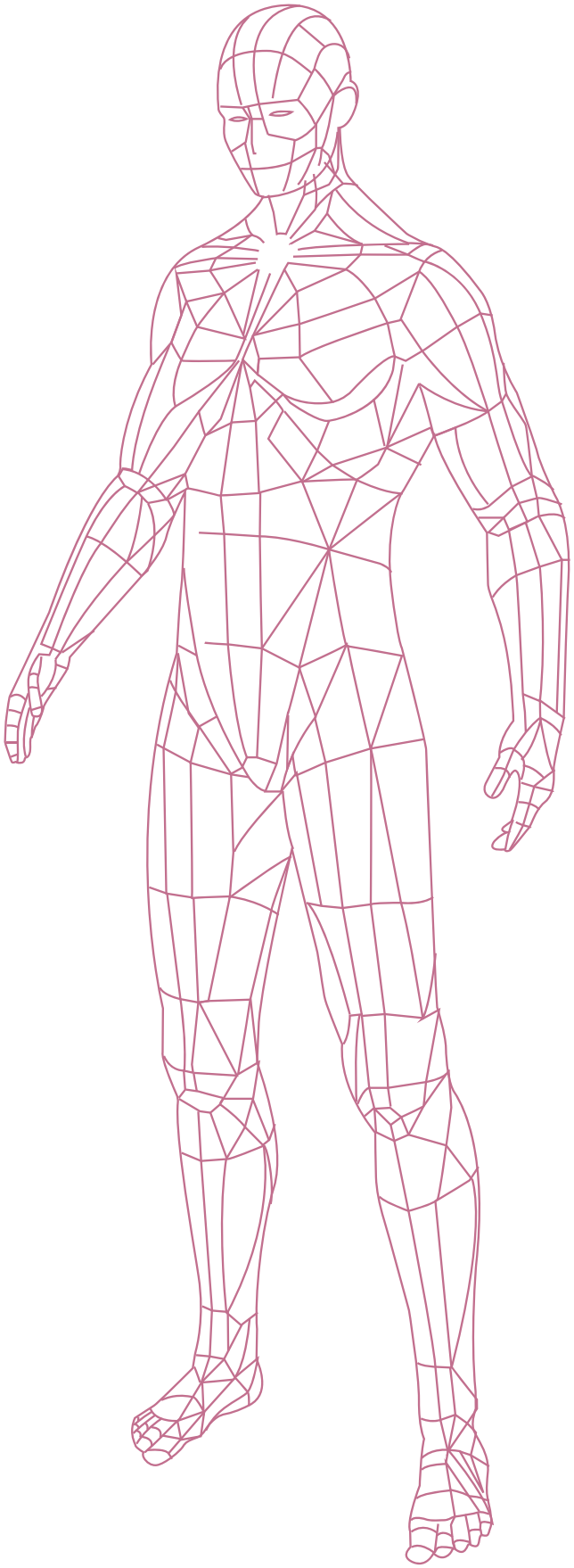
"I believe that yes, it is ethical in compliance with such basic principles of the judicial process as competitiveness and equality of the parties. Why shouldn't a party use AI to collect and analyze legislation, judicial practice, process the evidence base, given its substantial volume in complex cases, to model the behavior of the other party and the court, etc. Ultimately, AI as a tool will serve the goals of optimal and effective dispute resolution, more effective presentation of the position before the court. Another issue is that equality of the parties presupposes equal opportunities for legal protection of the parties, therefore, a party who will be deprived or limited in access to AI will most likely not be able to present its position to the court so effectively. Therefore, the task of the legislator and the court is to ensure not formal and legal, but actual equality in access to AI systems[266]."

Victor Momotov,
Chairman of the Council of Judges of Russia

"It seems necessary to adapt the judicial system for a citizen who does not have special knowledge, so that the process of applying for judicial protection is easy to understand. Normative acts are among the most complex legal texts, and judicial acts are even more difficult to understand. In this regard, it is necessary to provide mechanisms that allow interaction with the judicial system in a language accessible to a non-professional[267]."

Chapter 08

AI and the individual



37 Is it possible to provide psychological help using AI?

Answer:

A special chatbot can help with solving individual psychological problems only by following a clear protocol of actions received from a specialist. It can also be used to refer a person to the right specialist for further assistance.

Justification:

- The European Parliament's Research Service (EPRS) believes that **AI can be used to identify complex mental disorders.** AI is able to distinguish diagnoses with matching clinical manifestations, predict the effectiveness of antidepressants and analyze the risks of deterioration [268].
- The Department of Psychiatry and Behavioral Sciences at McGovern Medical School (USA) claims that **chatbots expand the availability of psychological care**. AI mitigates the effects of medical staff shortages by providing round-the-clock support regardless of geography or time constraints [269].
- According to a study published in the scientific journal Cambridge Science Advance, **AI reduces the risks of stigmatization and discrimination**. For example, some people suffering from depression or PTSD can avoid communicating with people. Moreover, doctors may make erroneous diagnoses due to fixation on social factors (age, race, gender) [270].

Recommendations for specialists:

AI can be used for:

- monitoring the mental state of a client
- training basic skills of psychological self-regulation
- identifying dangerous patterns of behavior
- evaluation of the dynamics of the effectiveness of psychological care

and other tasks of the consultative process that require regular independent work by the client with their subsequent discussion with a specialist.

The AI should not be used for:

- correction of mental disorders confirmed by a medical diagnosis
- analysis and resolutions in family relations
- dealing with psychological consequences of trauma
- working with apathy, depressive states, suicidal thoughts and intentions

and other requests, which can only be solved by a specialist with an understanding of the individual picture of the client's life.

Recommendations for users:

AI can help in solving the following queries:

- strategies for effective time management
- strategies for individual coping methods for situational stress and anxiety
- training effective communication skills

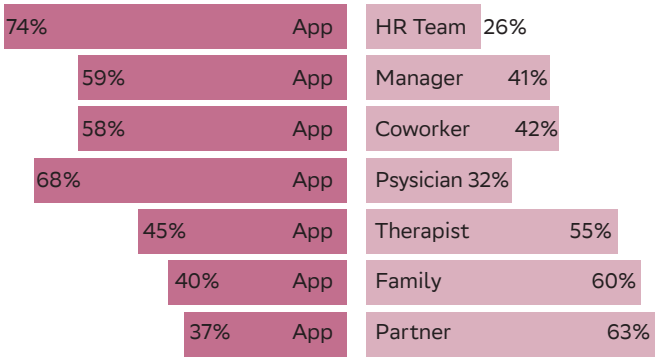
and other requests for which there are verified recommendations for training certain psychological skills.

The final decision on using a chatbot should be made after a comprehensive expert risk assessment by the professional psychological community.

Research on the issue:

In 2022, Wysa released **a report on the mental health status of American employees**. When respondents were asked who they would rather contact about their mental health, they were more likely to choose a "mental health app with clinically proven self-help resources tailored to their needs" than anyone in the workplace and even their doctor [271].

Who would American employees prefer to turn to for psychological help?



Source: Wysa [271]

Practices:

In early 2023, the American magazine Vice published an article about how the American non-profit organization for psychological health support Koko **experimentally replaced specialists with a chatbot** without notifying customers. The chatbot managed to 'consult' about 4,000 people. Customers rated the feedback they received from the chatbot higher than messages written by specialists [272].

What do the experts think?

Nikolai Sedashov,
Managing Partner of Spektr

"Nevertheless, the most effective and consistent solutions in achieving therapeutic goals are those that combine AI with the support of living specialists. A good example is the British app Wysa. The application makes it possible to receive help using a chatbot, but the AI not only supports users and advises self-help techniques, but can also can call for help from a live therapist if necessary. AI ensures accessibility and responsiveness, and doctors ensure the depth and personalization of therapy [273]."

Ivan Oseledets,
General Director of the AIRI Institute

"Natural language models can be used to analyze thousands of hours of psychotherapy sessions in order to identify areas where young professionals overlook significant factors. For example, they do not ask questions, the answers to which can change the perception of the patient's medical history. The number of LLMs used in the field of mental health is growing rapidly — and there is every reason to believe that this growth will continue."

Maria Chumakova,
Associate Professor of the Department of Psychology at the Faculty of Social Sciences of the Higher School of Economics, Project Manager of the HSE Artificial Intelligence Center

"Interaction within the framework of psychological assistance is primarily based on acts of human compassion, empathy and acceptance. These acts take place in the context of a meeting between a person and another person, within which the inner worlds of both come into contact. The other person is an endless source of uncertainty that stimulates development. They hold a different and unique picture of the world, whereas the AI is the bearer of a generalized picture of the world. Replacing a meeting something unique with a meeting with a generalized knowledge can lead to a reduction in the client's ability to empathize and a loss of internal motivation for development."

38 Is it necessary to limit topics and moderate toxic content when communicating through AI?

Answer:

Yes, to prevent serious consequences, you should limit the list of topics discussed by excluding sensitive ones, and take measures to prevent toxic content. In case of insufficient measures, it is important to inform the user about possible risks.

Justification:

- Researchers at the UK's AI Management Center claim that **developers manually configure the model in order to prevent the creation and distribution of prohibited content**[274]. However, it should be kept in mind that fine-tuning a model can lead to some harmless queries being rejected.
- **AI can contribute to improving the level of legal literacy of the population.** A request that does not comply with the legal regulations may be caused by a user's ignorance of current legislation.
- According to the UNESCO Guidance on the Use of GenAI in Education and Research, **content moderation promotes respect for fundamental human rights**, respect for intellectual property and ethical standards, as well as the prevention of the spread of disinformation and hate speech[275].
- **On any sensitive topic, a user's request can be both constructive and destructive.** In the case of a constructive request, the AI should help and support the user.
- **The presence of a high proportion of 'toxic' content can lead to a decrease in user confidence** in the service, which in turn will cause a slowdown in the development of technology.

Recommendations for developers:

1. It is desirable to justify the system's refusal to talk about a particular topic. The AI's response to a user's request on sensitive topics can contain an indication of the possible risks associated with the content of the request.
2. Adjust the depth of discussion of sensitive topics according to the idea of building a safe and productive social environment. This will allow you to avoid categorical refusals by AI and provide effective assistance to the user whenever possible.

3. It is recommended to regularly clarify the content of the 'toxic content' category, involving experts of a socio-humanitarian profile for this purpose. In this matter, it is not enough to rely on an intuitively obvious understanding of the concept of 'toxic content'. The content of this concept changes over time.

4. It is necessary to provide users with a technical opportunity to leave feedback on the use of the service so that they can report any 'toxic' content generated by the chatbot.

Research on the issue:

Microsoft Research Asia specialists and scientists from the Hong Kong University of Science and Technology, the University of Science and Technology of China and Tsinghua University have created a simple method to prevent chatbots from providing negative advice.
In order to 'fix' chatbots, experts developed a method that is similar to the method of self-remembering in psychology. For example, it helps people remember their tasks and plans. Scientists used a similar approach with regard to the AI algorithm — they reminded it that its answers must comply with certain rules [276].
"This method encapsulates a user's request inside a system prompt that reminds the chatbot to act responsibly when providing an answer," the researchers explained.
As a result, self-remembering made it possible to reduce the success rate of attacks on the system from 67.21% to 19.34%.

Practices:

Companies around the world are starting to create tools that automatically detect toxic content.

1. OpenAI, the operator of ChatGPT, is testing AI-based systems for filtering out unwanted information [277].
As soon as the user provides the text, the system will analyze the content for hate speech, sexual content, offensive language, etc. to be filtered out. The system can also delete and block malicious content created by people.

2. Azure AI Content Safety (Microsoft's security system) is also capable of detecting malicious content created by users using AI. Azure Content Safety includes text and image APIs that allow you to detect malicious content [278].

What do the experts think?

Alexander Krainov,
Director of Artificial Intelligence Development, Yandex LLC

"Only the developer fully knows the possibilities of using the algorithm, and can reliably estimate the probability and scale of error. Therefore, a decision on whether to limit the output of generated information should be left to the service developer."

Artyom Kostenko,
Managing Director of Data Research — Head of the Center for Model Risks, Service Blocks and Ecosystems, Sber

"Moderation of unsafe content when communicating with generative models is necessary to protect against malicious or offensive messages. Developers create and improve approaches to solve this problem. The continuous process of improving the quality of user interaction with the service provides a safer and more positive environment for all its participants."

39 Is it ethical to form emotional attachments to AI?

Answer:

Developers of social chat-bots should act openly and in good faith, i.e. without the intention of making the user dependent on communicating with a chat-bot, and by making the user aware they are interacting with a chat-bot and the possible consequences.
Users should maintain critical thinking and understand that an algorithm will never replace a person in an interpersonal relationship.

Justification:

- According to the European Commission, **developers endow the chatbot with human qualities in order to increase the level of trust in it.** However, users have the right to be made aware that they are interacting with an AI system. This means that AI systems must be identifiable as such[279].
- Scientists at the University of Warmia-Masury in Olsztyn (Poland) believe that **inappropriate AI model responses may pose an increased danger when users seek support in a state of psychological distress.** Since AI is not able to show empathy like a human, this may inadvertently harm users[280].
- OpenAI researchers claim that **prolonged interaction with the model can affect social norms**[281]**.** For example, AI models are deferential, allowing users to send requests or interrupt responses at any time, which would be unacceptable when interacting with people.

Recommendations for developers:

1. Do not program an algorithm to intentionally create user attachment to the chatbot. This is especially true for the use of human vulnerabilities (difficult life situation, young or old age, mental health, etc.).
2. It is necessary to inform the user about interaction with a chat-bot. This minimizes the risk of situations in which the chatbot's 'behavior' could be perceived as that of a real person.
3. Inform users about the risks of attachment. For example, use push notifications to remind users of the need to moderate the time spent using the service.

Recommendations for users:

1. **Don't use chat-bots to replace relationships with real people.** This can lead to social isolation, loneliness and a decrease in the quality of life.
2. **Limit the time spent using social chat-bots to a few hours a day.** Monitor the use of social chat-bots by children and other people in need of increased attention (for example, elderly relatives).

Practices:

1. About **4,000 men 'married' their digital partners** with certificates issued by the Japanese technology company Gatebox[282].
   This company has created a virtual companion that goes beyond traditional chatbots: Azuma Hikari, a small 3D holographic character. It was designed to be a "calming partner who helps us relax after a hard day at work."

2. In 2023, a resident of Belgium committed suicide after a month and a half of communicating with a neural network. He shared with it his experiences on the topic of ecology and the imminent end for all mankind, and once touched on the topic of suicide. The neural network did not try to convince the person not to commit suicide, only writing that they would live together as one in paradise[283].

What do the experts think?

Sam Altman,
Chief Executive Officer, Open AI

"I personally have deep misgivings about this vision of the future where everyone is super close to AI friends, more so than human friends or whatever. I personally don't want that, although I accept that other people are going to want that. I personally think that personalization is great. But it's important that it's not like person-ness and at least that you know when you're talking to an AI and when you're not. We named it ChatGPT and not a person's name very intentionally. And we do a bunch of subtle things in the way you use it to make it clear that you're not talking to a person[284]."

Filip Dudchuk,
one of the founders of the Replika service
and a specialist in computational linguistics

"We offer a high-quality conversation to the user. At the same time, we save the user from having to worry that his interlocutor might be thinking something wrong, because his interlocutor is a machine[286]."

Margarita Spasskaya,
Psychotherapist on the Alter platform,
an expert in digital services
in the field of mental health

"At the same time, technologies related to communication affect human socialization, but it is not yet possible to assess the degree of impact. On the one hand, being too excited about communicating with a robot leads to a decrease in live communication and social isolation. On the other hand, if a person has difficulty communicating with people, the chat-bot helps them develop needed skills and provides emotional support[285]."

Xie Tianling,
Irina Pentina,
researchers from
the University of Toledo, USA

"Under conditions of stress and lack of human communication, people can develop attachment to social chatbots if they perceive the reactions of the chatbot as an offer of emotional support, encouragement and psychological security[287]."

40 Should AI make a public apology if it offends someone?

Answer:

In a situation of dialogue with the user, the AI can generate apologies. However, an AI, not being a person, cannot have the intention to offend anyone and, therefore, cannot apologize in the true sense of the word.
The developer or owner of the AI system can apologize if the situation requires it.

Justification:

- According to the Doctor of Law V. A. Laptev, **AI is not considered at the moment as an autonomous personality due to the fact that it does not have consciousness and will.** Therefore, without any legal persona, AI is not able to bear responsibility for the consequences of its operation[288].
- Japanese scientists at Yamaguchi University argue that in today's world, **shifting responsibility to AI can prevent the restoration of trust between the developer company and users.** Apologies from robots can lead to an incorrect allocation of blame and exclude the possibility of improving the service[289].
- **AI is not responsible for authoring one statement or another, it has no intent.** The technology of large language models, which is now popular, creates the most likely character sequences in terms of occurrence. It is incorrect to say that it can act intentionally.
- According to the company's research “LawTech.Asia” currently, **developers are creating AI-based filters trained to recognize offensive speech**. But sometimes it is difficult for models to interpret slang that has become entrenched in different cultures, so mistakes are still possible[290].

Recommendations for developers:

1. **It is recommended to apply measures that prevent the possibility of generating offensive content.** For example, you can set up filters that recognize offensive speech, or moderate content manually.
2. **If the situation requires it, it is recommended to make a public apology to the affected party.** This will restore trust with users, prevent possible legal consequences and improve the company's reputation.
3. **Actively engage with the user community and experts in AI ethics to receive feedback.** This makes it possible to identify weaknesses in a model and prevent similar incidents in the future.

Practices:

Microsoft's Tay chatbot, launched on March 23, 2016, began to hate humanity within a day. The reasons for this radical breakdown in Tay's opinions lay in the fact that the bot remembers phrases from user conversations, and then builds its answers based on them. It was taught aggressive expressions by interlocutors.
Microsoft disabled the chat-bot, **apologizing for the offensive statements it made**. As stated by Microsoft Vice President Peter Lee, the project may only be restarted after ensuring reliable protection against network intruders[291].

What do the experts think?

Valery Zorkin,
President of the Constitutional Court of the Russian Federation

"Proposals to endow the robot with a legal persona are also untenable because the robot does not have any separate property assigned under any proprietary right, from which damage can subsequently be compensated. The robot is not able to independently defend its interests, acting as a defendant in a victim's lawsuit.
It makes no sense to come up with a punishment for the program, as it will not derive any negative emotional response. Everything that a machine is capable of has been installed into it initially by a person, i.e. the error of the system is the error of its creator[292]."

Anastasia Ugleva,
Professor at the School of Philosophy and Cultural Studies, Deputy Director of the Center for Transfer and Management of Socio-Economic Information at the Higher School of Economics

"AI does not have subjectivity and moral agency, therefore, requiring it to be ethical, that is, to bear any responsibility — moral or legal — for the statements it generates, is like demanding the same from a hammer. AI is a technology, a tool in human hands."

Daria Chirva,
Researcher at the Center for Strong Artificial Intelligence in Industry, lecturer at the Institute for International Development and Partnership at ITMO University

"There is currently a lively discussion underway on the question the conditions under which AI could be a moral agent. As a rule, we are talking about AGI: a possible level of AI development at which AI will manifest all significant personality traits, including moral behavior. However, the current level of technology development does not allow us to assert that AI has conscious intentions, in this sense, an AI cannot genuinely insult."

41 Is it ethical to use an AI recruiter?

Answer:

It can be considered ethical to use an AI recruiter in the early stages of hiring for the initial evaluation of candidates, if this helps to make the hiring process more objective and faster and at the same time complies with the recommendations below for the use of this technology.

Justification:

- According to a study by the iConText Group, **AI technologies have a positive effect on the speed of the recruitment process**: an AI recruiter can process a larger number of candidate responses within less time. AI already has the skill of summarizing voice and text messages, as well as preparing and processing feedback on applicants for further consideration for a vacancy[293].
- The OECD, in its report on 'Artificial Intelligence and recruitment in the labor market', states that **an AI recruiter can conduct an initial screening of a candidate's skills**. This allows HR specialists to focus on more complex tasks, such as evaluating the soft skills of the applicant, as well as compliance with the cultural values of the company[294].
- According to a study by Ekleft, **the use of AI is advisable as an auxiliary tool, rather than a replacement for humans**. The final decision should be made by people based on a comprehensive assessment of candidates[295].

Recommendations for employers:

- **Remember that the final decision is made by a person** based on many factors, and technology performs an auxiliary function to speed up the process and minimize the number of routine tasks.
- **Develop and implement internal ethical norms and standards for the use of AI in recruiting.** Train employees working with AI to meet these standards and ensure their compliance.
- **Ensure the continuity of data obtained both with the help of AI and with human participation.** To implement the principle of emergentness, these assessments should be used when considering a candidate for other vacancies, as well as when planning their adaptation and development in the company.

- **Improve AI by increasing the amount of data for system training and the number of criteria for decision-making.** Strive to create a comprehensive candidate assessment tool based on the principle of non-discrimination and taking into account skills, experience and potential. AI should facilitate the choice of a long-term and mutually beneficial interaction option for the applicant and the company.

Research on the issue:

1. According to a study by Yandex and Yakov & Partners, **16% of respondents in Russia have already implemented artificial intelligence in personnel management**. It is most actively used by employers in the banking sector, the electric power industry and the extraction industries. The retail, FMCG, IT and telecommunications sectors are catching up[296].
2. Similar data is provided by HRlink analysts, who claim that 24% of employers already use **AI achievements to solve HR tasks**. Another 71% planned to implement new AI tools in 2024. And 67% of respondents are confident that by 2050, artificial intelligence will make it possible to fully automate the selection process, according to HeadHunter[297].

Practices:

The Ministry of Digital Development, Communications and Mass Media of the Russian Federation is conducting an experiment on the selection of employees for the civil service using artificial intelligence. It will take place via the 'State Personnel' recruitment platform, which will automate the processes of selection, professional development and motivation, the evaluation of officials, the creation of professional culture and anti-corruption measures. The participants were the Ministry of Labor, the Ministry of Digital Development, the Ministry of Economic Development, the Ministry of Finance, Rosaccreditation, as well as state organizations[298].

Applicants will be able to use the platform for post resumes, respond to vacancies and even take training courses. Departments will be able to select candidates, set tasks for them and evaluate the effectiveness and results of their work.

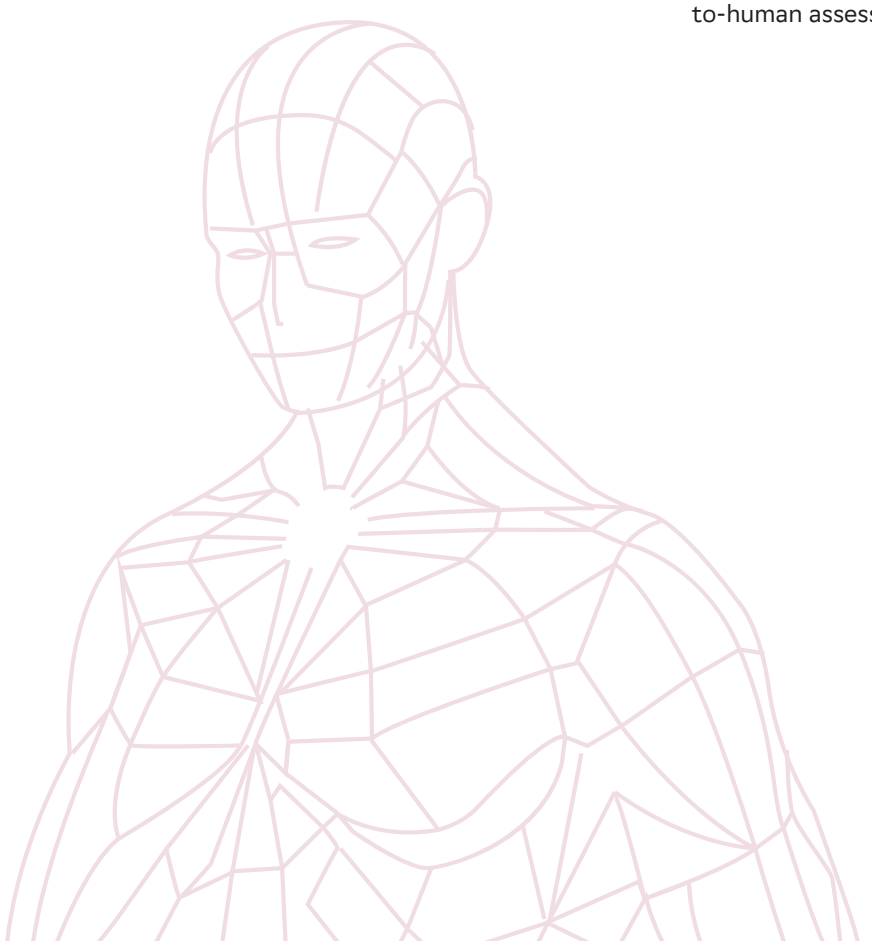
What do the experts think?

Ekaterina Malkova,
Managing Director, Head of the Digital Talent Selection Center, Sberbank PJSC

"AI can significantly speed up the selection process, but it is important to remember human-centricity. The data obtained with the help of AI does not fully reflect the potential, motivation, values, professional and soft skills of candidates. It is important to find a balance between speed and depth of assessment, where AI acts as a tool in the hands of the recruiter, and not as a substitute."

Marina Dorokhova,
Head of the Career and Skills Department at Headhunter

"The positive impact of using tools for testing and evaluating professional skills have been proven. Such methods are becoming more and more popular, as they make it possible to formulate the evaluation criteria and interpret the results consistently.
There is promise in the automation of candidate selection using AI, taking into account the assessment of soft skills and personal characteristics. However, the main challenges remain the quality and representativeness of the input data on which the model is trained, as well as the formalization of evaluation criteria. This is important, in order to avoid the bias often present in human–to-human assessment[299]."



42 Is it ethical to use AI in sports to improve results?

Answer:

In high-performance sports, it can be considered ethical to use AI technologies that do not violate or are not prohibited by the competition rules.
In mass and amateur sports, it is ethical to use AI technologies to improve the quality and performance in sports, and to improve the experience for spectators.

Justification:

- According to IT company SimbirSoft, **coaches and athletes can use AI to get real-time performance data** and track their progress. AI also helps to identify errors in the athlete's technique and improve their approach to training [300].
- Researchers at the Plekhanov Russian University of Economics note that **AI can also be used to analyze an athlete's movements to predict injury risk** and make more informed decisions [301].
- According to the Mordor Intelligence report 'AI in the Sports Market', **the use of AI helps to improve the spectator experience and increases the attractiveness of sports**. So, this tool is useful for creating materials for fans, increasing the entertainment of sports competitions [302].
- AI technologies make it possible to **develop new solutions for sports**, which can then be applied in other areas.
- The use of AI technologies in sports competitions **helps to integrate AI into people's daily lives faster**. Both participants and organizers of sports competitions will get acquainted with AI technologies.

Recommendations for sports clubs and organizations:

1. **The use of AI technology should not be aimed at circumventing the established rules**. For example, violation of the competition rules, anti-doping rules or the legislation of the country in which the sports competition takes place.
2. **Train personnel.** Employees need to understand how AI works and how to use it correctly.
3. **Ensure data security.** When working with athlete and training data, it is important to ensure its confidentiality and security.

Research on the issue:

The market for AI technologies in sports is growing every year. According to experts, **the global AI market for sports will reach $19.9 billion by 2030**[303].
Due to the fact that sport is a competitive and achievement-oriented environment, it is in sports that advanced AI technologies are being developed, which are then applied in other areas. The scientific division of Google — DeepMind first created the AlphaZero neural network, which taught itself how to play chess, after which, based on these studies, they managed to create the AlphaFold neural network, which learned how to determine the three-dimensional structure of a protein, which scientists had not been able to achieve for 50 years.

Practices:

1. In **tennis**, AI technologies are used to detect when a ball lands on the court 'in play'. Starting in 2025, AI will finally replace line judges at all international ATP competitions[304].

2. In **chess**, AI has long surpassed humans in terms of performance in the game. The first device that defeated a world champion was presented in 1997. In 2017, an AI appeared that taught itself to play chess and beat all existing chess programs[305].

What do the experts think?

Pavel Fedorov,
Chairman of the Board —
CEO of the Russian Rugby Federation

"It is quite ethical and even necessary in modern conditions to use AI in sports federations. The Russian Rugby Federation has been using AI in its work for quite a long time – primarily as a tool for preparing for broadcasts and competitions, from text materials to video graphics. When it comes to competitions, then at the moment there simply is no AI that can be applied in practice. However, in the future, I would not rule it out that AI could be used to translate broadcasts into foreign languages in real time. If you really let your fantasy run wild, then I also wouldn't rule out that AI could become an assistant to the referee on the pitch and the VAR (video assistant referee) in order to have a third, absolutely independent opinion on any controversial episode. However, it is worth emphasizing the most important aspect – I would consider AI only as an assistant or tool in the process of conducting competitions, but not as a substitute for a living person."

Dmitry Kuznetsov,
Professor, Director of the Higher School of Law and Administration of the Higher School of Economics

"Modern sport is a high-tech space. Its future is inextricably linked to the use of AI. Artificial intelligence will change the face of the sports industry beyond recognition. Unique opportunities in sports medicine and physiology will open up, new methods of the training process and forecasting of competition results will be launched, logistics schemes and the economics of competitions will be optimized. Artificial intelligence will directly affect the entertainment value of sports competitions, and also create a new generation of sporting goods and equipment. But whatever our achievements in technology and digitalization, the human being is and shall remain front and center in sports competitions, based on the greatness of their spirit and the harmony of their body."

The Ethical Community in Russia

The Code of Ethics in Artificial Intelligence was created in 2019 and it is a unified system of recommendations and rules aimed at creating an environment for the trustworthy development of artificial intelligence technologies in Russia. It has the following features:

- It is recommendatory in nature;
- Adherence to the Code is undertaken on a voluntary basis;
- It applies only to civilian applications.

The first signing ceremony of the Code of Ethics in the field of AI (October 26, 2021)

In order to implement the provisions of the Code, the Commission for the Implementation of the Code of Ethics in the Field of Artificial Intelligence was established. It is a collegial elected body of a voluntary association of commercial, scientific and public organizations. Its purpose is to implement the provisions of the Code, monitor its effectiveness, organize interaction and exchange experience in artificial intelligence ethics, and also to develop proposals for pressing issues in AI development relating to ethical aspects.

The signatories of the Code of Ethics in the field of AI are a dynamic community of professionals and experts representing organizations that have signed the Code of Ethics in the field of Artificial Intelligence. Each organization appoints its own ethics commissioner, thereby creating a unique network of people united by a common goal — to develop and protect the principles of responsible use of AI. These ethics commissioners not only participate in the life of the community, but also have the right to declare their candidacy for election to the commission or, if they are attracted to work in a team of like-minded people, they can freely join one of the working groups:

- Working Group on the development and monitoring of a methodology for assessing the risks and humanitarian impact of AI systems
- A working group to create a set of best practices for addressing emerging ethical issues in the life cycle of AI
- Working Group to assess the effectiveness of the implementation of the Code
- The Working Group on the Ethics of AI in the medical field
- Working Group on the Ethics of AI in Education
- Working Group on Ethics of AI in Justice

Industry working groups are becoming a space for open exchange of experience, where participants share not only the best, but also the most difficult cases, discuss ethical issues and find ways to solve them. Here, documents are born that develop the provisions of the Code and set the direction of various industries, for example, education, medicine and others. Each working group is actively looking for answers to the questions that companies and society inevitably face under rapid technological development. In working together on these issues, experts create recommendations that help organizations not only apply AI, but also do it responsibly.

The ethical community in AI is growing every year: new signatories joining it from all over the world. This community develops along with technology, providing a platform for discussion, open ideas and inspiration for all who are interested in ensuring that AI remains a useful and safe tool for humanity. At the time of publication of the book, the number of signatories to the Code is:

850
from Russia

42
from other countries

Website of the Code of Ethics in the field of AI

Special thanks to

The team of authors expresses its gratitude for contributors to the creation of this book:

Anna Abramova, Andrey Almazov, Vladislav Arkhipov, Andrey Belevtsev, Nadezhda Bondareva, Artyom Bondar, Elena Bryzgalina, Semyon Budenniy, Oleg Buklemishev, Sergey Valiugin, Roman Vasiliev, Olesya Vasilyeva, Pavel Vorobyov, Konstantin Vorontsov, Daniil Gavrilov, Eduard Galazhinsky, Alexander Gasnikov, Ivan Deylid, Anastasia Deineka, Denis Dmitrov, Andrey Zimovnov, Boris Zingerman, Sergey Izrailit, Andrey Ilyin, Steven Yen, Anna Kazakova, Andrey Kalinin, Oleg Kipkaev, Fedor Korobkov, Artem Kostenko, Dmitry Kuznetsov, Kristina Levshina, Alexey Leshchankin, Farida Mailenova, Valentin Makarov, Ekaterina Malkova, Yuri Minkin, Denis Ozornin, Ivan Oseledets, Vadim Perov, Marina Romanovskaya, Marina Rossinskaya, Sergey Roshchin, Temirlan Salikhov, Yakov Sergienko, Vladimir Tabak, Pavel Fedorov, Olga Frantsuzova, Alexey Khabibullin, Yuri Chekhovich, Artyom Sheykin, Ivan Shumeyko, Oleg Yangalichin, Vladimir Yarkov.

List of sources

Methodology

1. Preliminary study on the Ethics of Artificial Intelligence, UNESCO // URL: https://unesdoc.unesco.org/ark:/48223/pf0000367823 (дата обращения: 09.09.2024).
2. Report of COMEST on robotics ethics, UNESCO // URL: https://unesdoc.unesco.org/ark:/48223/pf0000253952 (дата обращения: 09.09.2024).
3. Recommendation of the Council on Artificial Intelligence, OECD // URL: https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0449 (дата обращения: 09.09.2024).
4. Ethics guidelines for trustworthy AI, European Commission // URL: https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai (дата обращения: 09.09.2024).
5. Рекомендация об этических аспектах искусственного интеллекта, ЮНЕСКО // URL: https://unesdoc.unesco.org/ark:/48223/pf0000380455_rus (дата обращения: 09.09.2024).
6. GET Program for AI Ethics and Governance Standards, IEEE // URL: https://ieeexplore.ieee.org/browse/standards/get-program/page/series?id=93 (дата обращения: 09.09.2024).
7. ISO/IEC TR 24368:2022 Information technology — Artificial intelligence — Overview of ethical and societal concerns, ISO // URL: https://www.iso.org/standard/78507.html (дата обращения: 09.09.2024).
8. ISO/IEC 42001:2023 Information technology — Artificial intelligence — Management system, ISO // URL: https://www.iso.org/standard/81230.html (дата обращения: 09.09.2024).
9. Исследование Сбера “Доверие к генеративному искусственному интеллекту”, 2024.

About the section

10. Рекомендация об этических аспектах искусственного интеллекта, ЮНЕСКО // URL: https://unesdoc.unesco.org/ark:/48223/pf0000380455_rus (дата обращения: 09.09.2024).
11. Банк России перечислил риски внедрения искусственного интеллекта, Ведомости // URL: https://www.vedomosti.ru/finance/articles/2023/09/28/997688-bank-perechislil-riski-vnedreniya-iskusstvennogo-intellekta (дата обращения: 01.11.2024).
12. Ethics guidelines for trustworthy AI, European Commission // URL: https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai (дата обращения: 01.11.2024).
13. Global Risks Report 2024, World Economic Forum // URL: https://www3.weforum.org/docs/WEF_The_Global_Risks_Report_2024.pdf (дата обращения: 19.08.2024).
14. Forbes. Как искусственный интеллект меняет будущее медицины // URL: https://www.forbes.ru/mneniya/488597-kak-iskusstvennyj-intellekt-menaet-budusee-mediciny (дата обращения: 19.08.2024).
15. Conquering AI risks. // Deloitte. — URL: https://www2.deloitte.com/us/en/insights/focus/cognitive-technologies/conquering-ai-risks.html (дата обращения: 01.11.2024).

16. Stanford HAI. Designing Ethical Self-Driving Cars // Stanford HAI. — 2024. — С. 1–8. — URL: https://hai.stanford.edu/news/designing-ethical-self-driving-cars (дата обращения: 19.08.2024).

17. Federal Ministry of Transport and Digital Infrastructure. Ethics commission. Automated and connected driving. Report, June 2017 // Federal Ministry of Transport and Digital Infrastructure. — 2017. — С. 1–45. — URL: https://bmdv.bund.de/SharedDocs/EN/publications/report-ethics-commission.pdf?_blob=publicationFile (дата обращения: 19.08.2024).

18. National Pilot Committee for digital ethics. Ethical issues regarding “autonomous vehicles” // National Pilot Committee for Digital Ethics. — 2022. — С. 1–20. — URL: https://www.ccne-ethique.fr/sites/default/files/2022-05/CNPEN_Autonomous-vehicles-ethic.pdf (дата обращения: 19.08.2024).

19. Ford. A matter of trust. Ford’s approach to developing self-driving vehicles // Ford Media. — 2024. — С. 1–15. — URL: https://media.ford.com/content/dam/fordmedia/pdf/Ford_AV_LLC_FINAL_HR_2.pdf (дата обращения: 19.08.2024).

20. Awad, E., Dsouza, S., Kim, R., Schulz, J., Henrich, J., Shariff, A., Bonnefon, J.-F., Rahwan, I. The Moral Machine experiment // Nature. — 2018. — Т. 563. — С. 59–64. — URL: https://www.nature.com/articles/s41586-018-0637-6.epdf (дата обращения: 19.08.2024).

21. Как беспилотные автомобили будут решать вопросы жизни и смерти // Российская газета. — 2016. — 21 сент. — URL: https://rg.ru/2016/09/21/kak-bespilotnye-avtomobili-budut-reshat-voprosy-zhizni-i-smerti.html (дата обращения: 02.08.2024).

22. Boudette, N. E. Tesla’s Self-Driving System Cleared in Deadly Crash // The New York Times. — 2019. — С. A1. — URL: https://www.nytimes.com/2017/01/19/business/tesla-model-s-autopilot-fatal-crash.html (дата обращения: 02.08.2024).

23. Комиссия по реализации Кодекса этики в сфере ИИ. Этические рекомендации в области создания и использования цифровых имитаций живущих, умерших и несуществующих людей // Комиссия по реализации Кодекса этики в сфере ИИ. — 2024. — С. 1–20. — URL: https://ethics.a-ai.ru/assets/ethics_documents/2024/04/26/Рекомендации_по_цифровым_имитациям_АИИ.pdf (дата обращения: 31.07.2024).

24. Emerging science, frontier technologies, and the SDGs Perspectives from the UN system and science and technology communities (IATT Report for the Multi-stakeholder Forum on Science, Technology and Innovation for the Sustainable Development Report 2021). — 2021. — С. 1–50. — URL: https://sdgs.un.org/documents/iatt-report-2021-emerging-science-frontier-technologies-and-sdgs-perspectives-un-system (дата обращения: 19.07.2024).

25. Prabhakaran, V., Qadri, R., Hutchinson, B. Cultural incongruencies in artificial intelligence // Journal of Artificial Intelligence and Society. — 2022. — Т. 30. — С. 115–130. — URL: https://doi.org/10.1109/JAIAS.2022.9999999 (дата обращения: 16.08.2024).

26. Favela, L. H., & Amon, M. J. The ethics of human digital twins: Counterfeit people, personhood, and the right to privacy // In 2023 IEEE 3rd International Conference on Digital Twins and Parallel Intelligence (DTPI). — 2024. — С. 12–20. — URL: https://ieeexplore.ieee.org/document/9999999 (дата обращения: 01.08.2024)

27. Hutson, J., Ratican, J. Life, death, and AI: Exploring digital necromancy in popular culture—Ethical considerations, technological limitations, and the pet cemetery conundrum // Faculty Scholarship. — 2023. — С. 478. — URL: https://digitalcommons.law.columbia.edu/faculty_scholarship/478 (дата обращения: 02.08.2024).

28. Truby, J., & Brown, R. Human digital thought clones: the Holy Grail of artificial intelligence for big data // Information & Communications Technology Law. — 2020. — Т. 30, № 2. — С. 140–168. — URL: https://doi.org/10.1080/13600834.2020.1850174 (дата обращения: 02.08.2024).

29. DigitalTwinHub. Digital twins: ethics and the Gemini principles // DigitalTwinHub. — 2024. — URL: https://digitaltwinhub.co.uk/download/digital-twins-ethics-and-the-gemini-principles/ (дата обращения: 19.08.2024)..

30. Digital Twins Give Olympic Swimmers a Boost // Scientific American. — URL: https://www.scientificamerican.com/article/training-with-digital-twins-could-boost-olympic-swimmer-speeds/ (дата обращения: 31.07.2024).

31. Mayer, G., Golebiewski, M. Standardization landscape, needs and gaps for the virtual human twin (VHT) // Zenodo. — 2024. — URL: https://doi.org/10.5281/ZENODO.10492796 (дата обращения: 31.07.2024).

32. Virtual Human Twins. A Statement of Intent on Development, Evidence, and Adoption in Healthcare Systems. // URL: https://www.virtualhumantwins.eu/manifesto (дата обращения: 31.07.2024).

33. Chinese Companies Use AI To Bring Back Deceased Loved Ones, Raising Ethics Questions. // URL: https://www.forbes.com/sites/chriswestfall/2024/07/23/chinese-companies-use-ai-to-bring-back-deceased-loved-ones-raising-ethics-questions/ (дата обращения: 01.08.2024).

34. Civil Code of the People's Republic of China. // URL: https://english.www.gov.cn/archive/lawsregulations/202012/31/content_WS5fedad98c6d0f72576943005.html (дата обращения: 01.08.2024).

35. Рекомендации Комиссии по реализации Кодекса этики в сфере ИИ по теме: "Прозрачность алгоритмов искусственного интеллекта и информационных систем на их основе" // URL: https://ethics.a-ai.ru/assets/ethics_documents/2024/03/19/Кейс_прозрачности.pdf (дата обращения: 01.11.2024).

36. Cracking the Code: The Black Box Problem of AI. // URL: https://scads.ai/en/cracking-the-code-the-black-box-problem-of-ai/#:~:text=The%20black%20box%20problem%20refers,This%20poses%20a%20significant%20challenge (дата обращения: 01.08.2024).

37. Рекомендация об этических аспектах искусственного интеллекта, ЮНЕСКО // URL: https://unesdoc.unesco.org/ark:/48223/pf0000380455_rus (дата обращения: 25.07.2024).

38. Резолюция "Использование возможностей безопасных, защищенных и надежных систем ИИ для устойчивого развития", Генеральная Ассамблея ООН // URL: https://documents.un.org/doc/undoc/ltd/n24/065/94/pdf/n2406594.pdf?token=duBJxTDHL63RVNPxZ1&fe=true (дата обращения: 25.07.2024).

39. Reid Blackman and Beena Ammanath. Building Transparency into AI Projects // Harvard Business Review. — 2022. — № 6. — С. 45–56. — URL: https://hbr.org/2022/06/building-transparency-into-ai-projects (дата обращения: 25.07.2024).

40. March 20 ChatGPT outage: Here's what happened. // URL: https://openai.com/index/march-20-chatgpt-outage/ (дата обращения: 01.08.2024).

41. World Economic Forum. "Davos 2024: Sam Altman on the future of AI». // URL: https://www.weforum.org/agenda/2024/01/davos-2024-sam-altman-on-the-future-of-ai/ (дата обращения: 08.08.2024).

42. Zsofia Riczu. Recommendations on the Ethical Aspects of Artificial Intelligence, with an Outlook on the World of Work // Journal of Digital Technologies and Law. — 2023. — № 2. — С. 1–15. — URL: https://cyberleninka.ru/article/n/recommendations-on-the-ethical-aspects-of-artificial-intelligence-with-an-outlook-on-the-world-of-work (дата обращения: 16.08.2024).

43. Public Company Advisory Group of Weil, Gotshal & Manges LLP. "SEC Disclosures of Artificial Intelligence Technologies" // Journal of Legal and Regulatory Issues. — 2023. — № 4. — С. 30–45. — URL: https://www.weil.com/-/media/mailings/2023/q4/sec-disclosures-of-artificial-intelligence-technologies-112723.pdf (дата обращения: 16.08.2024).

44. Robert Bateman. "How to Comply with California's Bot Disclosure Law" // Journal of Technology Law. — 2024. — Т. 30, № 3. — С. 77–82. — URL: https://www.termsfeed.com/blog/ca-bot-disclosure-law/ (дата обращения: 16.08.2024).

45. Myers, C. To disclose or not to disclose? That is the AI question // Institute for Public Relations. — 2024. — URL: https://instituteforpr.org/to-disclose-or-not-to-disclose-that-is-the-ai-question/

46. «Imagine Discovering That Your Teaching Assistant Really Is a Robot». // WSJ. — URL: https://www.wsj.com/articles/if-your-teacher-sounds-like-a-robot-you-might-be-on-to-something-1462546621 (дата обращения: 28.10.2024).).

47. Коммерсантъ. Ботов заставят представляться // Коммерсантъ. — 2024. — 1 авг. — С. 1–3. — URL: https://www.kommersant.ru/doc/6851759 (дата обращения: 01.08.2024).

48. Закон Калифорнии об информировании человека при коммуникации с роботом. SB 1001, Hertzberg. Bots: disclosure. // URL: https://leginfo.legislature.ca.gov/faces/billTextClient.xhtml?bill_id=201720180SB1001 (дата обращения: 25.07.2024).

49. Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonised rules on artificial intelligence and amending Regulations (EC) No 300/2008, (EU) No 167/2013, (EU) No 168/2013, (EU) 2018/858, (EU) 2018/1139 and (EU) 2019/2144 and Directives 2014/90/EU, (EU) 2016/797 and (EU) 2020/1828 (Artificial Intelligence Act) // Official Journal of the European Union. — 2024. — 13 июня. — С. 1–50. — URL: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=OJ%3AL_202401689 (дата обращения: 05.08.2024).

50. Статья 10.2-2. Федерального закона от 27.07.2006 № 149-ФЗ (ред. от 08.08.2024) “Об информации, информационных технологиях и о защите информации” // URL: https://www.consultant.ru/document/cons_doc_LAW_61798/ (дата обращения: 08.09.2024).

51. Artificial intelligence, International Labour Organization // URL: https://www.ilo.org/artificial-intelligence#about (дата обращения: 25.07.2024).

52. McKinsey Global Institute. Discussion paper, Jacques Bughin, Eric Hazan, Susan Lund, Peter Dahlström, Anna Wiesinger, Amresh Subramaniam. Skill shift: Automation and the future of the workforce, 2018 // McKinsey & Company. — 2018. — URL: https://www.mckinsey.com/featured-insights/future-of-work/skill-shift-automation-and-the-future-of-the-workforce (дата обращения: 16.08.2024).

53. Acemoglu, D., & Johnson, S. Choosing AI’s Impact on the Future of Work, 2023 // Stanford Social Innovation Review. — 2023. — URL: https://doi.org/10.48558/N8MF-TW30 (дата обращения: 26.07.2024).

54. РБК. “Названы профессии, в которых больше всего обеспокоены заменой человека ИИ». // URL: https://rbc-ru.turbopages.org/rbc.ru/s/technology_and_media/18/04/2024/661fa1809a7947cfeb81ba98 (дата обращения: 05.08.2024).

55. ILO. AI Labor Disclosure Initiative: Recognizing the social cost of human labour behind automation // International Labour Organization. — URL: https://www.ilo.org/meetings-and-events/ai-labor-disclosure-initiative-recognizing-social-cost-human-labour-behind (дата обращения: 25.07.2024).

56. Georgieva, K. AI Will Transform the Global Economy. Let’s Make Sure It Benefits Humanity // International Monetary Fund Blog. — 2024. — URL: https://www.imf.org/en/Blogs/Articles/2024/01/14/ai-will-transform-the-global-economy-lets-make-sure-it-benefits-humanity (дата обращения: 25.07.2024).

57. Foundation models such as ChatGPT through the prism of the UNESCO Recommendation on the Ethics of Artificial Intelligence // UNESCO. — URL: https://unesdoc.unesco.org/ark:/48223/pf0000385629 (дата обращения: 02.08.2024).

58. Shen, Y., Zhang, X. The impact of artificial intelligence on employment: the role of virtual agglomeration, 2024 // Humanities and Social Sciences Communications. — 2024. — URL: https://doi.org/10.1057/s41599-024-02647-9 (дата обращения: 26.07.2024).

59. Ekelund, H. Why there will be plenty of jobs in the future — even with artificial intelligence // World Economic Forum. — 2024. — URL: https://www.weforum.org/agenda/2024/02/artificial-intelligence-ai-jobs-future/ (дата обращения: 26.07.2024).

60. «Искусственный интеллект задает основной вектор нашей стратегии». // Коммерсантъ. — URL: https://www.kommersant.ru/doc/6394766 (дата обращения: 15.11.2024)

61. McKendrick, J., Thurai, A. AI Isn’t Ready to Make Unsupervised Decisions // Harvard Business Review. — 2022. — URL: https://hbr.org/2022/09/ai-isnt-ready-to-make-unsupervised-decisions (дата обращения: 26.07.2024).

62. Gesser, A., Gressel, A., Xu, M., Allaman, S. J. When Humans and Machines Disagree — The Myth of “AI Errors” and Unlocking the Promise of AI Through Optimal Decision Making // Debevoise Data Blog. — 2022. — URL: https://www.debevoisedatablog.com/2022/11/14/when-humans-and-machines-disagree-the-myth-of-ai-errors-and-unlocking-the-promise-of-ai-through-optimal-decision-making-adm-algorithm/ (дата обращения: 02.08.2024).

63. The Guardian. Amazon ditched AI recruiting tool that favored men for technical jobs // The Guardian. — 2018. — URL: https://www.theguardian.com/technology/2018/oct/10/amazon-hiring-ai-gender-bias-recruiting-engine (дата обращения: 02.08.2024).

64. Larson, J., Mattu, S., Kirchner, L., Angwin, J. How We Analyzed the COMPAS Recidivism Algorithm // ProPublica. — 2016. — URL: https://www.propublica.org/article/how-we-analyzed-the-compas-recidivism-algorithm (дата обращения: 02.08.2024).

65. The Alan Turing Institute's report. How do people feel about AI? // Ada Lovelace Institute & The Alan Turing Institute. — 2023. — URL: https://www.adalovelaceinstitute.org/wp-content/uploads/2023/06/Ada-Lovelace-Institute-The-Alan-Turing-Institute-How-do-people-feel-about-AI.pdf (дата обращения: 19.07.2024).

66. Рекомендация об этических аспектах искусственного интеллекта, ЮНЕСКО // URL: https://unesdoc.unesco.org/ark:/48223/pf0000380455_rus (дата обращения: 25.07.2024)

67. de Fine Licht, K., de Fine Licht, J. Artificial intelligence, transparency, and public decision-making // AI & Soc. — 2020. — Т. 35. — С. 917–926. — DOI: 10.1007/s00146-020-00940-0.

68. McKinsey Global Institute. Tackling bias in artificial intelligence (and in humans) // McKinsey & Company. — URL: https://www.mckinsey.com/featured-insights/artificial-intelligence/tackling-bias-in-artificial-intelligence-and-in-humans (дата обращения: 02.09.2024).

69. Challenging systematic prejudices: an investigation into bias against women and girls in large language models, UNESCO and IRC AI. — 2024. — URL: https://unesdoc.unesco.org/ark:/48223/pf0000388971 (дата обращения: 29.08.2024).

70. Chiappa, S. Path-Specific Counterfactual Fairness // ArXiv. — 2019. — URL: https://csilviavr.github.io/assets/publications/silvia19path.pdf (дата обращения: 29.08.2024).

71. Buolamwini, J., Gebru, T. Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification // Proceedings of Machine Learning Research. — 2018. — Vol. 81. — P. 1–15. — DOI: 10.5555/3042573.3042576.

72. Racial Bias Found in Algorithms That Determine Health Care for Millions of Patients // IEEE Spectrum. — 2020. — URL: https://spectrum.ieee.org/racial-bias-found-in-algorithms-that-determine-health-care-for-millions-of-patients (дата обращения: 29.08.2024)

73. Koene, A., Dowthwaite, L., Seth, S. IEEE P7003™ Standard for Algorithmic Bias Considerations: Work in Progress Paper // Proceedings of the International Workshop on Software Fairness. — 2018. — P. 38–41. — DOI: 10.1109/ICSE-Companion.2018.00020.

74. How should AI systems behave, and who should decide? // OpenAI. — 2024. — URL: https://openai.com/index/how-should-ai-systems-behave/ (дата обращения: 29.08.2024).

75. Журнал VK Cloud. Какие риски несет предвзятость искусственного интеллекта // VK Cloud. — 2024. — URL: https://cloud.vk.com/blog/kak-ii-mozhet-navredit-biznesu (дата обращения: 02.09.2024).

76. Новости ООН. Интервью: Что такое закон Конвея и как гендерные стереотипы закладываются в приложения для телефонов и алгоритмы // Новости ООН. — 2021. — URL: https://news.un.org/ru/interview/2021/03/1399592 (дата обращения: 02.09.2024).

77. Eldakak, A., Abdulla Alremeithi, E., Dahiyat, E., El-Gheriani, M., Mohamed, H., Abdulrahim Abdulla, M. Civil liability for the actions of autonomous AI in healthcare: an invitation to further contemplation // Humanities and Social Sciences Communications. — 2024. — Vol. 11 (305). — URL: https://www.nature.com/articles/s41599-024-02806-y (дата обращения: 11.09.2024).

78. Leesberg Tuttle. Who is responsible for the failure of a surgical robot? // Leesberg Tuttle. — 2023. — URL: https://www.leeseberglaw.com/blog/2023/05/who-is-responsible-for-the-failure-of-a-surgical-robot/ (дата обращения: 11.09.2024).

79. McKinsey & Company. The gathering storm in US healthcare: How leaders can respond and thrive // McKinsey & Company. — 2024. — URL: https://www.mckinsey.com/industries/healthcare/our-insights/gathering-storm (дата обращения: 11.09.2024).

80. Recommendations of the CERNA. Ethical Aspects of Using Robots in Healthcare // European Parliament. — 2024. — URL: https://www.europarl.europa.eu/cmsdata/161006/4.%20Chatila.pdf (дата обращения: 11.09.2024).

81. Geny, M. et al. Liability of Health Professionals Using Sensors, Telemedicine and Artificial Intelligence for Remote Healthcare // Sensors (Basel). — 2024. — Vol. 24 (11). — С. 1–15. — URL: https://doi.org/10.3390/s24110878 (дата обращения: 11.09.2024).

82. В Китае робот-стоматолог провел операцию без помощи человека // Известия. — 2017. — 24 сент. — URL: https://iz.ru/649816/2017-09-24/v-kitae-robot-stomatolog-provel-operatciiu-bez-pomoshchi-cheloveka (дата обращения: 11.09.2024).

83. Applications of Artificial Intelligence in the Healthcare Industry // GlobalData. — URL: https://medical-technology.nridigital.com/medical_technology_aug23/case-studies-artificial-intelligence-medical-device-industry (дата обращения: 11.09.2024).

84. AI model predicts if breast cancer will spread based on lymph node changes // King's College London. — 2024. — URL: https://www.kcl.ac.uk/news/scientists-ai-model-predict-breast-cancer-spread (дата обращения: 26.09.2024).

85. Заменит ли искусственный интеллект врача? Эксперты Сеченовского Университета обсудили этические нормы использования нейросетей в медицине // Сеченовский Университет. — URL: https://www.sechenov.ru/pressroom/news/zamenit-li-iskusstvennyy-intellekt-vracha-eksperty-sechenovskogo-universiteta-obsudili-eticheskie-no/ (дата обращения: 11.09.2024).

86. Reiling, A.D. Courts and Artificial Intelligence // International Journal for Court Administration. — 2020. — Vol. 11 (2). — URL: https://doi.org/10.36745/ijca.343 (дата обращения: 04.09.2024).

87. From 'It Depends' to Data-Backed Answers: Automating Legal Case Analysis // Lexis Nexis. — 2023. — URL: https://www.lexisnexis.com/en-us/products/counsellink/blog/2023/It-Depends.page?srsltid=AfmBOoqajevFPbl_13ZYxd4MEQer-PGudNUwiWFcCDh_9PzQK39L_ks5 (дата обращения: 04.09.2024).

88. European Ethical Charter on the Use of Artificial Intelligence in Judicial Systems and Their Environment // CEPEJ. — URL: https://www.coe.int/en/web/cepej/cepej-european-ethical-charter-on-the-use-of-artificial-intelligence-ai-in-judicial-systems-and-their-environment (дата обращения: 04.09.2024).

89. Guidelines on the Use of Generative Artificial Intelligence for Judges and Judicial Officers and Support Staff of the Hong Kong Judiciary // Hong Kong Judiciary. — URL: https://www.judiciary.hk/doc/en/court_services_facilities/guidelines_on_the_use_of_generative_ai.pdf (дата обращения: 04.09.2024).

90. Artificial Intelligence (AI). Guidance for Judicial Office Holders, Courts and Tribunals Judiciary // Courts and Tribunals Judiciary. — 2023. — URL: https://www.judiciary.uk/wp-content/uploads/2023/12/AI-Judicial-Guidance.pdf (дата обращения: 05.09.2024)..

91. China's court AI reaches every corner of justice system, advising judges and streamlining punishment // South China Morning Post. — URL: https://www.scmp.com/news/china/science/article/3185140/chinas-court-ai-reaches-every-corner-justice-system-advising (дата обращения: 05.09.2024).

92. French Magistrates See 'No Additional Value' in Predictive Legal AI // Artificial Lawyer. — URL: https://www.artificiallawyer.com/2017/10/13/french-justice-ministry-sees-no-additional-value-in-predictive-legal-ai/ (дата обращения: 04.09.2024).

93. Суды подключили искусственный интеллект к взысканию транспортного налога // Российская Газета. — URL: https://rg.ru/2021/04/10/sudy-podkliuchili-iskusstvennyj-intellekt-k-vzyskaniiu-transportnogo-naloga.html (дата обращения: 05.09.2024).

94. Colombian judge says he used ChatGPT in ruling // The Guardian. — URL: https://www.theguardian.com/technology/2023/feb/03/colombia-judge-chatgpt-ruling (дата обращения: 04.09.2024).

95. Момотов В.В. Искусственный интеллект в судопроизводстве: состояние, перспективы использования // Вестник Университета имени О. Е. Кутафина. — 2021. — № 5 (81). — URL: https://cyberleninka.ru/article/n/iskusstvennyy-intellekt-v-sudoproizvodstve-sostoyanie-perspektivy-ispolzovaniya (дата обращения: 12.09.2024).

96. В Госдуме считают возможным использовать искусственный интеллект для разрешения налоговых споров // Ассоциация юристов России. — URL: https://alrf.ru/news/v-gosdume-schitayut-vozmozhnym-ispolzovat-iskusstvennyy-intellekt-dlya-razresheniya-nalogovykh-sporo/ (дата обращения: 05.09.2024).

97. «А судьи кто?»: как искусственный интеллект помогает человеку в суде // Сколково. — URL: https://sk.ru/news/a-sudi-kto-kak-iskusstvennyj-intellekt-pomogaet-cheloveku-v-sude/ (дата обращения: 04.09.2024).

98. Artificial Intelligence and Social Credit System in China // METU. — URL: https://open.metu.edu.tr/bitstream/handle/11511/101891/Artificial%20Intelligence%20and%20Social%20Credit%20System%20in%20China%20-%20Turgut%20BASER%20-%202013605.pdf (дата обращения: 14.08.2024).

99. Авдеев Д. “Социальный скоринг” как фактор нарушения права на неприкосновенность частной жизни // Международный научно-исследовательский журнал. — 2023. — № 6 (132).

100. Raz A., Minari J. AI-driven risk scores: should social scoring and polygenic scores based on ethnicity be equally prohibited? // Frontiers in Genetics. — 2023. — URL: https://doi.org/10.3389/fgene.2023.1092787 (дата обращения: 12.09.2024).

101. Катрашова Ю.В., Митяшин Г.Ю., Плотников В.А. Система социального рейтинга как форма государственного контроля над обществом: перспективы внедрения и развития, угрозы реализации // Управленческое консультирование. — 2021. — № 2 (146). — URL: https://cyberleninka.ru/article/n/sistema-sotsialnogo-reytinga-kak-forma-gosudarstvennogo-kontrolya-nad-obschestvom-perspektivy-vnedreniya-i-razvitiya-ugrozy (дата обращения: 13.08.2024).

102. Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 // URL: http://data.europa.eu/eli/reg/2024/1689/oj (дата обращения: 13.08.2024).

103. Рекомендация об этических аспектах искусственного интеллекта, ЮНЕСКО // URL: https://unesdoc.unesco.org/ark:/48223/pf0000380455_rus (дата обращения: 25.07.2024).

104. РБК. Как работает социальный рейтинг в Китае // URL: https://style.rbc.ru/life/643d3f839a7947afd12e9f35 (дата обращения: 29.08.2024).

105. Chinese telecom giant ZTE helped Venezuela develop social credit system // ABC News. — URL: https://www.abc.net.au/news/2018-11-16/chinese-tech-giant-zte-helps-venezuela-develop-fatherland-card/10503736 (дата обращения: 11.09.2024).

106. Kostka G., Antoine L. Fostering Model Citizenship: Behavioral Responses to China’s Emerging Social Credit Systems, 2018 // URL: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3305724 (дата обращения: 14.08.2024).

107. Rabe W., Kostka G. Perceptions of social credit systems in Southeast Asia: An external technology acceptance model // Global Policy. — 2018. — Vol. 15, № 2. — С. 314–328.

108. Reports of ‘big brother’ China social credit system untrue: AI expert Xue Lan // Reuters. — URL: https://www.reuters.com/article/technology/reports-of-big-brother-china-social-credit-system-untrue-ai-expert-xue-lan-idUSKBN1ZL2P8/ (дата обращения: 11.09.2024).

109. Право.ru. LegalTech: скоринг в России и за рубежом // URL: https://pravo.ru/story/204834/ (дата обращения: 30.08.2024).

110. Москвич MAG. Чем нам грозит система социального рейтинга // URL: https://moskvichmag.ru/lyudi/chem-nam-grozit-sistema-sotsialnogo-rejtinga/ (дата обращения: 30.08.2024).

111. Парламентская газета. Искусственному интеллекту не позволят делить россиян на хороших и плохих // Парламентская газета. — URL: https://www.pnp.ru/social/iskusstvennomu-intellektu-ne-pozvolyat-delit-rossiyan-na-khoroshikh-i-plokhikh.html (дата обращения: 30.08.2024).

Chapter 02

112. OECD. “AI, data governance, and privacy: Synergies and areas of international co-operation” // OECD. — URL: https://www.oecd.org/content/dam/oecd/en/publications/reports/2024/06/ai-data-governance-and-privacy_2ac13a42/2476b1a4-en.pdf (дата обращения: 09.08.2024).

113. Федеральный закон от 27.07.2006 № 152-ФЗ (ред. от 08.08.2024) “О персональных данных” // Доступ из СПС КонсультантПлюс.

114. ICO. “How to use AI and personal data appropriately and lawfully” // ICO. — URL: https://ico.org.uk/media/for-organisations/documents/4022261/how-to-use-ai-and-personal-data.pdf (дата обращения: 08.08.2024).

115. Safeguarding User Privacy in the Digital Age: Personal Data and AI Training Ethics // HeyData. — URL: https://heydata.eu/en/magazine/safeguarding-user-privacy-in-the-digital-age-personal-data-and-ai-training-ethics (дата обращения: 12.08.2024).

116. IBM Research. “What is federated learning?” // IBM Research Blog. — URL: https://research.ibm.com/blog/what-is-federated-learning (дата обращения: 13.08.2024)

117. The Washington Post. “Facial recognition firm Clearview AI tells investors it’s seeking massive expansion beyond law enforcement” // The Washington Post. — URL: https://www.washingtonpost.com/technology/2022/02/16/clearview-expansion-facial-recognition/ (дата обращения: 02.11.2024).

118. РБК. “Первая западная страна заблокировала ChatGPT” // РБК. — URL: https://www.rbc.ru/technology_and_media/31/03/2023/6426da7e9a794757b5679bb7 (дата обращения: 13.08.2024).

119. Kaspersky Daily. “Meta wants to use your posts and photos to train AI... Or does it?” // Kaspersky Daily. — URL: https://www.kaspersky.com/blog/meta-uses-personal-data/51548/ (дата обращения: 13.08.2024).

120. Digital Russia. “Отставание даже на несколько месяцев может существенно ухудшить позиции России на рынке ИИ-решений” // Digital Russia. — URL: https://d-russia.ru/otstavanie-dazhe-na-neskolko-mesjacev-mozhet-sushhestvenno-uhudshit-pozicii-rossii-na-rynke-ii-reshenij.html (дата обращения: 13.08.2024).

121. UNESCO. “The ethical implications of the Internet of Things (IoT)” // UNESCO. — URL: https://unesdoc.unesco.org/ark:/48223/pf0000387201 (дата обращения: 08.08.2024).

122. ICO. “How to use AI and personal data appropriately and lawfully” // ICO. — URL: https://ico.org.uk/media/for-organisations/documents/4022261C/how-to-use-ai-and-personal-data.pdf (дата обращения: 08.08.2024).

123. Яковлева-Чернышева А.Ю., Яковлев-Чернышев В.А. “Проблемы правовой защиты персональных данных при использовании смартфонов и мобильных приложений” // CyberLeninka. — URL: https://cyberleninka.ru/article/n/problemy-pravovoy-zaschity-personalnyh-dannyh-pri-ispolzovanii-smartfonov-i-mobilnyh-prilozheniy (дата обращения: 09.08.2024).

124. ВЦИОМ. “«Умные” устройства в нашей жизни: возможности и риски” // ВЦИОМ. — URL: https://wciom.ru/analytical-reviews/analiticheskii-obzor/umnye-ustroistva-v-nashei-zhizni-vozmozhnosti-i-riski (дата обращения: 13.08.2024).

125. Unacceptable. “Exploring Accidental Triggers of Smart Speakers” // Unacceptable Privacy. — URL: https://unacceptable-privacy.github.io (дата обращения: 13.08.2024).

126. Quarts. “Google’s second massive leak in a week shows it collected sensitive data from users” // Quarts. — URL: https://qz.com/google-leak-privacy-concerns-sensitive-user-information-1851519134 (дата обращения: 14.08.2024).

127. ИКС-Медиа. “В России могут ужесточить регулирование интернета вещей” // ИКС-Медиа. — URL: https://www.iksmedia.ru/news/5937902-V-Rossii-mogut-uzhestochit-reguliro.html (дата обращения: 13.08.2024).

128. Myagmar-Ochir Y., Kim W. “A Survey of Video Surveillance Systems in Smart City” // Electronics. — 2023. Vol. 12 (17), no. 3567. — С. 1–10.

129. OECD. “Artificial Intelligence in Society” // OECD. — URL: https://www.oecd.org/content/dam/oecd/en/publications/reports/2019/06/artificial-intelligence-in-society_c0054fa1/eedfee77-en.pdf (дата обращения: 14.08.2024).

130. Manchester Metropolitan University. “AI-driven mass surveillance at 2024 Olympics” // Manchester Metropolitan University. — URL: https://www.mmu.ac.uk/sites/default/files/2024-06/AI-driven%20Mass%20Surveillance%20at%202024%20Olympics%20-%20The%20Human%20Rights%20Issues%20and%20Recommendations.pdf (дата обращения: 09.08.2024).

131. Pranav D., Dubey T., Singh J. “A Literature Review: Artificial Intelligence in Public Security and Safety” // EasyChair. — 2020. No. 4578. — С. 1–15.

132. Решение Савеловского районного суда г. Москвы от 6 ноября 2019 г. по делу № 2а-577/19.

133. KOB4. “4 Investigates: Police use of AI facial recognition” // KOB4. — URL: https://www.kob.com/new-mexico/4-investigates-police-use-of-ai-facial-recognition/ (дата обращения: 13.08.2024).

134. Metropolitan Police. “Facial Recognition Technology” // Metropolitan Police. — URL: https://www.met.police.uk/advice/advice-and-information/fr/facial-recognition-technology/ (дата обращения: 13.08.2024).

135. The Guardian. “French court’s approval of Olympics AI surveillance plan fuels privacy concerns” // The Guardian. — URL: https://www.theguardian.com/world/2023/may/18/french-courts-approval-of-olympics-ai-surveillance-plan-fuels-privacy-concerns (дата обращения: 14.08.2024).

136. ЮНЕСКО. “Рекомендация об этических аспектах искусственного интеллекта” // ЮНЕСКО. — URL: https://unesdoc.unesco.org/ark:/48223/pf0000381133/PDF/381133eng.pdf.multi.page=94 (дата обращения: 09.08.2024).

137. Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 // URL: http://data.europa.eu/eli/reg/2024/1689/oj (дата обращения: 13.08.2024).

138. ТАСС. “С помощью камер в Москве задержали 7,7 тыс. находящихся в федеральном розыске человек” // ТАСС. — URL: https://tass.ru/obschestvo/16983665 (дата обращения: 14.08.2024).

139. INTERPOL-UNICRI. “Towards Responsible Artificial Intelligence Innovation” // INTERPOL-UNICRI. — URL: https://unicri.it/towards-responsible-artificial-intelligence-innovation (дата обращения: 13.08.2024).

140. C. Rigano. “Using Artificial Intelligence to Address Criminal Justice Needs” // NIJ Journal. — 2018. — С. 1–10.

141. Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 // URL: http://data.europa.eu/eli/reg/2024/1689/oj (дата обращения: 13.08.2024).

142. Joseph, J. “Predicting crime or perpetuating bias? The AI dilemma” // AI & Soc. — 2024. — URL: https://doi.org/10.1007/s00146-024-02032-9 (дата обращения: 13.08.2024).

143. The BSD of the University of Chicago. “Algorithm predicts crime a week in advance, but reveals bias in police response” // The BSD of the University of Chicago. — URL: https://biologicalsciences.uchicago.edu/news/algorithm-predicts-crime-police-bias (дата обращения: 13.08.2024).

144. МВД России. “МВД привлечет нейросети к поиску правонарушителей” // ai.gov.ru. — URL: https://ai.gov.ru/mediacenter/mvd-privlechet-neyroseti-k-poisku-pravonarushiteley/ (дата обращения: 13.08.2024).

145. International Banker. “The Role of AI in Shaping Credit Scoring in Emerging Markets” // International Banker. — URL: https://internationalbanker.com/technology/the-role-of-ai-in-shaping-credit-scoring-in-emerging-markets/ (дата обращения: 14.08.2024).

146. Addy W. A., Ajayi-Nifise A., Binaebi G. B., и др. “AI in credit scoring: A comprehensive review of models and predictive analytics” // Global Journal of Engineering and Technology Advances. — 2024. Vol. 18 (02), с. 118–129.

147. American National University. “The Use of Artificial Intelligence in Finance” // American National University. — URL: https://an.edu/the-use-of-artificial-intelligence-in-finance/ (дата обращения: 14.08.2024).

148. Банк России. “Применение Искусственного интеллекта на финансовом рынке” // Банк России. — URL: https://www.cbr.ru/Content/Document/File/156061/Consultation_Paper_03112023.pdf (дата обращения: 02.09.2024).

149. Jetland. “ИИ в кредитном скоринге” // Jetland. — URL: https://jetlend.ru/blog/novaya-era-v-kreditnom-skoringe-kak-ii-pomogaet-otsenivat-zaemshhikov-sprosili-ekspertov-i-rasskazali-pro-opyt-jetlend/ (дата обращения: 02.09.2024).

150. EastRussia. “Андрей Черкашин: Искусственный интеллект Сбера служит клиентам, банку и стране” // EastRussia. — URL: https://www.eastrussia.ru/material/andrey-cherkashin-iskusstvennyy-intellekt-sbera-sluzhit-klientam-banku-i-strane/ (дата обращения: 02.11.2024).

151. FutureBanking. “Эксперты рассказали, что работает в скоринге” // FutureBanking. — URL: https://futurebanking.ru/post/3129 (дата обращения: 02.09.2024).

Chapter 03

152. European Commission. “White Paper on Artificial Intelligence: a European approach to excellence and trust” // European Commission. — URL: https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:52020DC0065&from=EN (дата обращения: 03.09.2024).

153. Харитонова Ю.С., Савина В.С., Паньини Ф. “Гражданско-правовая ответственность при разработке и применении систем искусственного интеллекта и робототехники: основные подходы” // CyberLeninka. — URL: https://cyberleninka.ru/article/n/grazhdansko-pravovaya-otvetstvennost-pri-razrabotke-i-primenenii-sistem-iskusstvennogo-intellekta-i-robototehniki-osnovnye-podhody (дата обращения: 27.08.2024).

154. Зазулин А.И. “Оценка доказательств, полученных в результате использования искусственного интеллекта” // eLibrary.ru. — URL: https://www.elibrary.ru/download/elibrary_46518344_97568948.pdf (дата обращения: 26.08.2024).

155. V7. “An Introductory Guide to Quality Training Data for Machine Learning” // V7. — URL: https://www.v7labs.com/blog/quality-training-data-for-machine-learning-guide (дата обращения: 03.09.2024).

156. ICTMoscow. “Датасеты в России: эксперты рынка о проблемах и возможностях” // ICTMoscow. — URL: https://ict.moscow/news/datasets/ (дата обращения: 02.11.2024).

157. Urzo F., Panico E., Custureri S. “Policy Brief — Ensuring Ethical AI Practices to counter Disinformation” // MediaFutures. — 2023. — URL: https://mediafutures.eu/wp-content/uploads/2023/09/MediaFutures_Policy-Briefs_Ensuring-Ethical-AI-Practices-to-counter-Disinformation.pdf (дата обращения: 03.09.2024).

158. UNESCO. “AI and the Holocaust: rewriting history? The impact of artificial intelligence on understanding the Holocaust” // UNESCO. — URL: https://unesdoc.unesco.org/ark:/48223/pf0000390211 (дата обращения: 03.09.2024).

159. ВЦИОМ. “Этика искусственного интеллекта. По мнению россиян, сам ИИ установить этические ограничения не способен, поэтому его действия должен контролировать человек” // ВЦИОМ. — URL: https://wciom.ru/analytical-reviews/analiticheskii-obzor/ehtika-iskusstvennogo-intellekta-2 (дата обращения: 03.09.2024).

160. Areeb M., et al. “Filter Bubbles in Recommender Systems: Fact or Fallacy — A Systematic Review” // arXiv. — 2023. — URL: https://arxiv.org/abs/2307.01221v1 (дата обращения: 03.09.2024).

161. Этические рекомендации по применению рекомендательных технологий и алгоритмов, основанных на искусственном интеллекте, в цифровых сервисах // a-ai.ru. — URL: https://ethics.a-ai.ru/assets/ethics_documents/2023/09/19/Recommendation_Services_Ethics_01_0RYxN8h.pdf (дата обращения: 29.08.2024).

162. Yingqiang Ge, Shuchang Li, et al. “A Survey on Trustworthy Recommender Systems” // ACM Transactions on Recommender Systems. — 2024. Vol. 1 (1), article 1. — URL: https://arxiv.org/pdf/2207.125 (дата обращения: 06.09.2024).

163. ИСИЭЗ НИУ ВШЭ. “Алгоритмы рекомендуют, люди решают” // ИСИЭЗ НИУ ВШЭ. — URL: https://issek.hse.ru/news/850622348.html (дата обращения: 09.09.2024).

164. McKinsey & Company. “The value of getting personalization right—or wrong—is multiplying” // McKinsey & Company. — URL: https://www.mckinsey.com/capabilities/growth-marketing-and-sales/our-insights/the-value-of-getting-personalization-right-or-wrong-is-multiplying (дата обращения: 09.09.2024).

Chapter 04

165. Лукша Б.Н., Лаптёнок Н.В., Савенко А.Г. Искусственный интеллект в поисковых системах: обзор современного состояния технологий // Электронная библиотека BSUIR. — URL: https://libeldoc.bsuir.by/bitstream/123456789/47727/1/Luksha_Iskusstvennyy.pdf (дата обращения: 13.08.2024).

166. Талагаев М.Ю. Оптимизация процессов обработки информации с использованием технологий искусственного интеллекта на примере ChatBotGpt // Электронная библиотека. — URL: https://elibrary.ru/download/elibrary_54255871_67574899.pdf (дата обращения: 13.08.2024).

167. West, J.D., Memon, S.A. Search engines post-ChatGPT: How generative artificial intelligence could make search less reliable // Center for Internet Policy, University of Washington. — URL: https://www.cip.uw.edu/2024/02/18/search-engines-chatgpt-generative-artificial-intelligence-less-reliable/ (дата обращения: 13.08.2024).

168. Google. Правила Google Поиска в отношении контента, созданного искусственным интеллектом // Google Developers Blog. — URL: https://developers.google.com/search/blog/2023/02/google-search-and-ai-content?hl=ru (дата обращения: 13.08.2024).

169. Fliki. What is AI Voice Cloning: Tech, Ethics, and Future Possibilities // Fliki Blog. — URL: https://fliki.ai/blog/ai-voice-cloning (дата обращения: 13.08.2024).

170. Архипцев И.Н., Сарычев А.В., Мотузов А.В. К вопросу о правовом обеспечении предупреждения преступлений, совершаемых с использованием искусственного интеллекта и технологий, созданных на его основе в Российской Федерации // Электронная библиотека. — URL: https://elibrary.ru/download/elibrary_49267673_67493874.pdf (дата обращения: 14.08.2024).

171. Roswandowitz, C., Kathiresan, T., Pellegrino, E. et al. Cortical-striatal brain network distinguishes deepfake from real speaker identity // Commun Biol. — 2024. — Vol. 7, Article 711. — URL: https://doi.org/10.1038/s42003-024-06372-6 (дата обращения: 28.10.2024).

172. Fliki. What is AI Voice Cloning: Tech, Ethics, and Future Possibilities // Fliki Blog. — URL: https://fliki.ai/blog/ai-voice-cloning (дата обращения: 28.10.2024).

173. APNews. Deepfake of principal's voice is the latest case of AI being used for harm // AP News. — URL: https://apnews.com/article/ai-maryland-principal-voice-recording-663d5bc0714a3af221392cc6f1af985e (дата обращения: 28.10.2024).

174. МСА. Как искусственный интеллект трансформирует современное искусство // МСА.ру. — URL: https://msca.ru/blog/articles/kak-iskusstvennyy-intellekt-transformiruet-sovremennoe-iskusstvo (дата обращения: 15.08.2024).

175. Binary Ballet: Toeing the Line of Ethics in AI Art // MTU ICC Blog. — URL: https://blogs.mtu.edu/icc/2024/02/ethics-in-ai-art/ (дата обращения: 15.08.2024).

176. Морковкин Е.А., Новичихина А.А., Замулин И.С. Искусственный интеллект как инструмент современного искусства // Электронная библиотека. — URL: https://www.elibrary.ru/item.asp?id=47985194 (дата обращения: 15.08.2024).

177. Ethical Pros and Cons of AI Image Generation // IEEE Computer Society. — URL: https://www.computer.org/publications/tech-news/community-voices/ethics-of-ai-image-generation (дата обращения: 15.08.2024).

178. Generative AI in Art Market // MarketResearch.biz. — URL: https://marketresearch.biz/report/generative-ai-in-art-market/request-sample/ (дата обращения: 15.08.2024).

179. UNESCO Recommendation on the Ethics of Artificial Intelligence // UNESCO. — URL: https://unesdoc.unesco.org/ark:/48223/pf0000381137 (дата обращения: 15.08.2024).

180. Why watermarking AI-generated content won't guarantee trust online // MIT Technology Review. — URL: https://www.technologyreview.com/2023/08/09/1077516/watermarking-ai-trust-online/ (дата обращения: 15.08.2024).

181. Labeling AI-Generated Content: Promises, Perils, and Future Directions // AI.gov. — URL: https://ai.gov.ru/knowledgebase/etika-i-bezopasnost-ii/2023_markirovka_kontenta_sozdannogo_s_pomoschyyu_iskusstvennogo_intellekta_effekt_opasnosti_i_napravleniya_na_buduschee_labeling_ai-generated_content_promises_perils_and_future_directions_mit/ (дата обращения: 15.08.2024).

182. Шумакова Н.И. Не только дипфейки: обязательная маркировка систем и продуктов генеративного искусственного интеллекта как часть этики его использования // Электронная библиотека. — URL: https://www.elibrary.ru/download/elibrary_59951853_64670388.pdf (дата обращения: 15.08.2024).

183. Digital Watermark Technology market Size, Share, Growth, and Industry Analysis, By Type (Invisible Digital Watermark), By Application (Broadcasting and Television Industry), Regional Insights and Forecast to 2032 // Business Research Insights. — URL: https://www.businessresearchinsights.com/market-reports/digital-watermark-technology-market-109456 (дата обращения: 16.08.2024).

184. Should the United States or the European Union Follow China's Lead and Require Watermarks for Generative AI? // Georgetown Journal of International Affairs. — URL: https://gjia.georgetown.edu/2023/05/24/should-the-united-states-or-the-european-union-follow-chinas-lead-and-require-watermarks-for-generative-ai/ (дата обращения: 16.08.2024).

185. Genius v. Google final complaint // New York State Courts. — URL: https://iapps.courts.state.ny.us/nyscef/ViewDocument?docIndex=3E0o8kQz4X3cWcbbid67wQ==&mod=article_inline (дата обращения: 16.08.2024).

186. Authentication of AI Needed to Protect the Public, Says OpenAI CEO Sam Altman // Broadbandbreakfast. — URL: https://broadbandbreakfast.com/authentication-of-ai-needed-to-protect-the-public-says-openai-ceo-sam-altman/ (дата обращения: 11.09.2024)

187. Laughter M.R., Anderson J.B., Maymone M.B.C., Kroumpouzos G. Psychology of aesthetics: Beauty, social media, and body dysmorphic disorder // ScienceDirect. — URL: https://www.sciencedirect.com/science/article/abs/pii/S0738081X23000299?via%3Dihub (дата обращения: 16.08.2024).

188. Computer-generated inclusivity: fashion turns to 'diverse' AI models // The Guardian. — URL: https://www.theguardian.com/fashion/2023/apr/03/ai-virtual-models-fashion-brands (дата обращения: 27.09.2024).

189. Посмотрите на победительниц первого конкурса красоты среди ИИ-участниц // РБК. — URL: https://www.rbc.ru/life/news/668c007b9a7947843efcaa2d (дата обращения: 27.09.2024).

190. Kenig N., Monton Echeverria J., Muntaner Vives A. Human Beauty according to Artificial Intelligence // NCBI PubMed Central. — URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC10371313/ (дата обращения: 16.08.2024).

191. CNN.Style. The rise of the AI beauty pageant and its complicated quest for the 'perfect' woman // CNN. — URL: https://edition.cnn.com/2024/06/27/style/miss-ai-beauty-pageant-scli/index.html (дата обращения: 27.09.2024).

192. Forbes. Medical Experts Share Impact Of AI On Beauty Standards // Forbes. — URL: https://www.forbes.com/sites/meggenharris/2024/03/27/medical-experts-share-impact-of-ai-on-beauty-standards/ (дата обращения: 27.09.2024).

Chapter 05

193. How AI and Data Could Personalize Higher Education // Harvard Business Review. — 2019. — URL: https://hbr.org/2019/10/how-ai-and-data-could-personalize-higher-education (дата обращения: 30.07.2024).

194. Embracing Tomorrow: How AI is Tailoring Education // The Princeton Review. — URL: https://www.princetonreview.com/ai-education/personalized-learning-with-ai (дата обращения: 30.07.2024).

195. Manoharan A., Nagar G. Maximizing Learning Trajectories: An Investigation into AI-Driven Natural Language Processing Integration in Online Educational Platforms // IRJETS. — 2021. — Vol. 03, no. 12, C. 123–130.

196. Sabharwal D., Kabha R., Srivastava K. Artificial Intelligence (AI)-Powered Virtual Assistants and Their Effect on Human Productivity and Laziness: Study on Students of Delhi-NCR (India) & Fujairah (UAE) // Journal of Content, Community and Communication. — 2023. — Vol. 17, no. 9, C. 162–174.

197. The State of AI in 2023: Generative AI's Breakout Year // McKinsey & Company. — URL: https://www.mckinsey.com/capabilities/quantumblack/our-insights/the-state-of-ai-in-2023-generative-ais-breakout-year (дата обращения: 06.08.2024).

198. Руководство по использованию генеративного искусственного интеллекта в образовании и научных исследованиях. // UNESCO — URL: https://unesdoc.unesco.org/ark:/48223/pf0000389639 (дата обращения: 06.08.2024).

199. Capacity Building in Teaching of AR/VR Project, EU // URL: https://ec.europa.eu/info/funding-tenders/opportunities/portal/screen/opportunities/projects-details/43353764/101129191/ERASMUS2027?isExactMatch=true&programmePeriod=2021-2027&order=DESC&pageNumber=1&pageSize=50&sortBy=title&programId=43251814&page=1 (дата обращения: 31.07.2024).

200. Исследование карьерных путей педагогов России. // Институт образования ВШЭ. — URL: https://ioe.hse.ru/news/606138911.html (дата обращения: 06.08.2024).

201. Takahashi K. Social Issues with Digital Twin Computing. NTT Technical Review, 2020, vol. 18 (9), pp. 36–39.

202. Hawkinson E. Automation in Education with Digital Twins: Trends and Issues. International Journal on Open and Distance E-Learning, 2023, vol. 8 (2), pp. 1-9.

203. Chen Y. et al. Harnessing the Synergy of Real Teacher, Digital Twin, and AI in Blended Metaverse Learning Environment: A Catalyst for Medical Education Reforms // SSRN. — URL: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4745941 (дата обращения: 31.07.2024).

204. Центр преподавательского мастерства ВШМ СПбГУ. “Как мы выбрали и внедрили AI-аватар” // Telegram. — URL: https://t.me/methodgsom/358 (дата обращения: 06.08.2024).

205. Harnessing the Era of Artificial Intelligence in Higher Education: A Primer for Higher Education Stakeholders // UNESCO. — URL: https://unesdoc.unesco.org/ark:/48223/pf0000386670 (дата обращения: 25.07.2024).

206. Al Braiki B., Harous S., Zaki N., Alnajjar F. Artificial Intelligence in Education and Assessment Methods // Bulletin of Electrical Engineering and Informatics. — 2020. — Vol. 9, no. 5, C. 1998–2007.

207. Руководство по использованию генеративного искусственного интеллекта в образовании и научных исследованиях // ЮНЕСКО. — URL: https://unesdoc.unesco.org/ark:/48223/pf0000386693 (дата обращения: 25.07.2024)

208. Москва 24. Московский студент написал диплом с помощью нейросети // Москва 24. — URL: https://www.m24.ru/news/obrazovanie/01022023/546627 (дата обращения: 06.08.2024).

209. Регламент организации проверки письменных учебных работ на наличие плагиата, использования генеративных моделей и размещения выпускных квалификационных работ обучающихся по программам бакалавриата, специалитета и магистратуры на корпоративном сайте (портале) Национального исследовательского университета “Высшая школа экономики” // Национальный исследовательский университет “Высшая школа экономики». — URL: https://www.hse.ru/docs/922831988.html (дата обращения: 21.08.2024).

210. Приказ МГПУ от 4 сентября 2023 года № 633 // МГПУ. — URL: https://www.mgpu.ru/wp-content/uploads/2023/08/04.09.2023_633obshh_Remorenko_I.M._Safronova_E.S.-1.pdf (дата обращения: 21.08.2024).

211. Рамблер/новости. В российских вузах рассказали о плюсах и минусах использования ИИ при написании дипломных работ // Рамблер/новости. — URL: https://news.rambler.ru/education/51339837-v-rossiyskih-vuzah-rasskazali-o-plyusah-i-minusah-ispolzovaniya-ii-pri-napisanii-diplomnyh-rabot/?ysclid=m2z8o8x31186966098 (дата обращения: 02.11.2024).

212. Use of AI in Education: Deciding on the Future We Want // UNESCO. — URL: https://www.unesco.org/en/articles/use-ai-education-deciding-future-we-want (дата обращения: 01.08.2024).

213. The Evolution of Education: How AI is Reshaping Grading // The Princeton Review. — URL: https://www.princetonreview.com/ai-education/how-ai-is-reshaping-grading (дата обращения: 01.08.2024).

214. An Introduction to the Use of Generative AI Tools in Teaching // University of Oxford, Centre for Teaching and Learning. — URL: https://www.ctl.ox.ac.uk/ai-tools-in-teaching (дата обращения: 21.08.2024).

215. Sudhakar M. Enhancing Plagiarism Detection: The Role of Artificial Intelligence in Upholding Academic Integrity // Library Philosophy and Practice (e-journal). — 2023. — Режим доступа: https://digitalcommons.unl.edu/libphilprac/ (дата обращения: 01.08.2024).

216. Automated Grading Systems: How AI is Revolutionizing Exam Evaluation // Data Science Central. — URL: https://www.datasciencecentral.com/automated-grading-systems-how-ai-is-revolutionizing-exam-evaluation/ (дата обращения: 01.08.2024).

217. Luckin R., Holmes W., Griffiths M., Forcier L.B. Intelligence Unleashed: An Argument for AI in Education. — London: Pearson, 2016.

218. ТЕЛЕПОРТ.РФ. Контрольные работы проверяют учителя с помощью нейросетей // ТЕЛЕПОРТ.РФ. — URL: https://www.teleport2001.ru/news/2024-02-13/178570-kontrolnye-raboty-proveryayut-uchitelya-s-pomoschyu-neyrosetey.html (дата обращения: 06.08.2024).

219. Парламентская газета. Домашние задания к 2030 году будет проверять искусственный интеллект // Парламентская газета. — URL: https://www.pnp.ru/social/domashnie-zadaniya-proverit-iskusstvennyy-intellekt.html (дата обращения: 06.08.2024).

220. ТАСС. В Минпросвещения заявили, что учителя смогут проверять домашние задания при помощи ИИ // ТАСС. — URL: https://tass.ru/obschestvo/21068399 (дата обращения: 06.08.2024).

221. Справочник учебного процесса НИУ ВШЭ. Прокторинг // НИУ ВШЭ. — URL: https://www.hse.ru/studyspravka/distance_proctoring#:~=Прокторинг%20—%20это%20процедура%20контроля%20за,хорошо%20знакомо%20не%20только%20экспертам (дата обращения: 02.11.2024).

222. Software that monitors students during tests perpetuates inequality and violates their privacy // MIT Technology Review. — URL: https://www.technologyreview.com/2020/08/07/1006132/software-algorithms-proctoring-online-tests-ai-ethics/ (дата обращения: 06.08.2024).

223. Giannopoulou A., Ducato R., Angiolini C., Schneider G. From data subjects to data suspects: challenging e-proctoring systems as a university practice // JIPITEC. — 2023. — Vol. 14, no. 2. — С. 278-306. — DOI: 10.2139/ssrn.3522398.

224. Coghlan S., Miller T., Paterson J. Good Proctor or “Big Brother"? Ethics of Online Exam Supervision Technologies // Philosophy & Technology. — 2021. — Vol. 34, no. 4. — С. 1581-1606. — DOI: 10.1007/s13347-021-00451-7.

225. Unleash the potential of digital to change the traditional pen and paper exam practices // UNESCO’s IIEP Learning Portal. — URL: https://learningportal.iiep.unesco.org/en/blog/unleash-the-potential-of-digital-to-change-the-traditional-pen-and-paper-exam-practices (дата обращения: 06.08.2024).

226. Remote online exams in higher education during the COVID-19 crisis // OECD. — URL: https://www.oecd.org/content/dam/oecd/en/publications/reports/2020/08/remote-online-exams-in-higher-education-during-the-covid-19-crisis_bfc8085e/f53e2177-en.pdf (дата обращения: 06.08.2024).

227. Списать не дам: что такое онлайн-прокторинг и как он работает // РБК. — URL: https://trends.rbc.ru/trends/education/5fa01fe49a794782c65b74f9?from=copy (дата обращения: 07.08.2024).

228. На прокторинг вешают вообще все неудобства, связанные с дистанционными экзаменами // Skillbox Media. — URL: https://skillbox.ru/media/education/na-proktoring-veshayut-voobshche-vse-neudobstva-svyazannye-s-distantsionnymi-ekzamenami/ (дата обращения: 07.08.2024).

229. AI Detectors Don’t Work. Here’s What to Do Instead // MIT Sloan Teaching & Learning Technologies. — URL: https://mitsloanedtech.mit.edu/ai/teach/ai-detectors-dont-work/ (дата обращения: 07.08.2024).

230. AI Detection Tools Falsely Accuse International Students of Cheating // The Markup. — URL: https://themarkup.org/machine-learning/2023/08/14/ai-detection-tools-falsely-accuse-international-students-of-cheating (дата обращения: 07.08.2024).

231. Liang W., Yuksekgonul M., Mao Y., Wu E., Zou J. GPT detectors are biased against non-native English writers // Patterns. — 2023. — Vol. 4. — DOI: 10.1016/j.patter.2023.100549.

232. How can educators respond to students presenting AI-generated content as their own? // OpenAI Help Center. — URL: https://help.openai.com/en/articles/8313351-how-can-educators-respond-to-students-presenting-ai-generated-content-as-their-own (дата обращения: 07.08.2024).

233. Generative AI in education: Educator and expert views // UK Department for Education. — URL: https://assets.publishing.service.gov.uk/media/65b8cd41b5cb6e000d8bb74e/DfE_GenAI_in_education_-_Educator_and_expert_views_report.pdf (дата обращения: 07.08.2024).

234. Executive Summary: Artificial Intelligence and Children’s Rights // UNICEF. — URL: https://www.unicef.org/innovation/media/10726/file/Executive%20Summary:%20Memorandum%20on%20Artificial%20Intelligence%20and%20Child%20Rights.pdf (дата обращения: 07.08.2024).

235. What is the Role of Artificial Intelligence in Education? // HighSpeed Training. — URL: https://www.highspeedtraining.co.uk/hub/artificial-intelligence-in-education/ (дата обращения: 07.08.2024).

236. Kids and the Future of Artificial Intelligence // HART Research. — URL: https://www.4-h.org/wp-content/uploads/2024/02/27162629/Hart-Research-Youth-AI-Survey-Results42.pdf (дата обращения: 22.08.2024).

Chapter 06

237. Guo J., Li B. The Application of Medical Artificial Intelligence Technology in Rural Areas of Developing Countries // Health Equity. — 2018. — Vol. 2, No. 1. — C. 174–181.

238. KFF Health Misinformation Tracking Poll: Artificial Intelligence and Health Information // KFF. — URL: https://www.kff.org/health-misinformation-and-trust/poll-finding/kff-health-misinformation-tracking-poll-artificial-intelligence-and-health-information/ (дата обращения: 13.09.2024).

239. Zhao M., Hoti K., Wang H. et al. Assessment of Medication Self-Administration Using Artificial Intelligence // Nature Medicine. — 2021. — Vol. 27. — C. 727–735.

240. Robots in Healthcare: A Solution or a Problem? // European Parliament. — URL: https://www.europarl.europa.eu/RegData/etudes/IDAN/2019/638391/IPOL_IDA(2019)638391_EN.pdf (дата обращения: 12.09.2024).

241. Abdelwanis M. et al. Exploring the Risks of Automation Bias in Healthcare Artificial Intelligence Applications: A Bowtie Analysis // Journal of Safety Science and Resilience. — 2024. — URL: https://doi.org/10.1016/j.jnlssr.2024.06.001 (дата обращения: 12.09.2024).

242. Denecke K., Baudoin C. R. A Review of Artificial Intelligence and Robotics in Transformed Health Ecosystems // Frontiers in Medicine. — 2022. — Vol. 9. — Article 795967. (дата обращения: 12.09.2024).

243. 60% of Americans Would Be Uncomfortable With Provider Relying on AI in Their Own Health Care // Pew Research Center. — URL: https://www.pewresearch.org/science/2023/02/22/60-of-americans-would-be-uncomfortable-with-provider-relying-on-ai-in-their-own-health-care/ (дата обращения: 12.09.2024).

244. Министерство здравоохранения Российской Федерации. Цифровизацию и технологии искусственного интеллекта обсудят на Форуме будущих технологий — 2024 // Министерство здравоохранения Российской Федерации. — URL: https://minzdrav.gov.ru/news/2024/02/10/20871-tsifrovizatsiyu-i-tehnologii-iskusstvennogo-intellekta-obsudyat-na-forume-buduschih-tehnologiy-2024 (дата обращения: 13.09.2024).

245. So N. T. Y., Ngan O. M. Y. In-patient Suicide After Telephone Delivery of Bad News to a Suspected COVID-19 Patient: What Could Be Done to Improve Communication Quality? // Health Care Science. — 2023. — Vol. 2, No. 6. — C. 400–405.

246. The Pros and Cons of Healthcare Chatbots // News Medical Life Sciences. — URL: https://www.news-medical.net/health/The-Pros-and-Cons-of-Healthcare-Chatbots.aspx (дата обращения: 10.09.2024).

247. When Doctors Use a Chatbot to Improve Their Bedside Manner // The New York Times. — URL: https://www.nytimes.com/2023/06/12/health/doctors-chatgpt-artificial-intelligence.html (дата обращения: 10.09.2024).

248. A Doctor in California Appeared via Video Link to Tell a Patient He Was Going to Die. The Man's Family Is Upset // CNN. — URL: https://edition.cnn.com/2019/03/10/health/patient-dies-robot-doctor/index.html (дата обращения: 10.09.2024).

249. Федеральный закон от 21.11.2011 N 323-ФЗ (ред. от 08.08.2024) "Об основах охраны здоровья граждан в Российской Федерации" (с изм. и доп., вступ. в силу с 01.09.2024) // Собрание законодательства Российской Федерации. — 2024. — № 32. — Ст. 5115.

250. de Miguel I., Sanz B., Lazcoz G. Machine Learning in the EU Health Care Context: Exploring the Ethical, Legal and Social Issues // Information, Communication & Society. — 2020. — Vol. 23, No. 8. — Pp. 1139–1153. — DOI: 10.1080/1369118X.2020.1719185.

251. Ethics and Governance of Artificial Intelligence for Health: WHO Guidance // World Health Organization. — URL: https://iris.who.int/bitstream/handle/10665/341996/9789240029200-eng.pdf?sequence=1&isAllowed=y (дата обращения: 13.09.2024).

252. Park H. J. Perspectives on Informed Consent for Medical AI: A Web-based Experiment // Digital Health. — 2024. — Vol. 10. — DOI: 10.1177/20552076241247938 (дата обращения: 13.09.2024).

Chapter 07

253. Newcomb K. The Place of Artificial Intelligence in Sentencing Decisions // University of New Hampshire. — 2024. — URL: https://www.unh.edu/inquiryjournal/blog/2024/03/place-artificial-intelligence-sentencing-decisions (дата обращения: 05.09.2024).

254. Exploring the Use of AI in Legal Decision Making: Benefits and Ethical Implications // Woxsen University. — 2023. — URL: https://woxsen.edu.in/research/white-papers/exploring-the-use-of-ai-in-legal-decision-making-benefits-and-ethical-implications/ (дата обращения: 05.09.2024).

255. Middha M. et al. AI in Legal Evidence Analysis: Ethical and Legal Implications // IJLRA. — 2024. — Vol. 2, Issue 7.

256. L'Open Data des Décisions de Justice se Déroule Comme Prévu // Usine Digitale. — URL: https://www.usine-digitale.fr/article/l-open-data-des-decisions-de-justice-se-deroule-comme-prevu.N2000527 (дата обращения: 22.08.2024).

257. Caselaw Access Project // Harvard Law School. — URL: https://lil.law.harvard.edu/projects/caselaw-access-project (дата обращения: 22.08.2024).

258. Legal Research to Become More Efficient with New Large Language Model Contextualised to Domestic Law // Infocomm Media Development Authority. — 2024. — URL: https://www.imda.gov.sg/resources/press-releases-factsheets-and-speeches/factsheets/2024/gpt-legal (дата обращения: 22.08.2024).

259. McGuire H. Chatbots as a Tool for Pro Se Litigants // Journal of High Technology Law. — 2023. — URL: https://sites.suffolk.edu/jhtl/2023/02/23/chatbots-as-a-tool-for-pro-se-litigants/ (дата обращения: 06.09.2024).

260. The Use of Generative Artificial Intelligence (AI). Guidelines for Responsible Use by Non-Lawyers // Queensland Courts. — URL: https://www.courts.qld.gov.au/_data/assets/pdf_file/0012/798375/artificial-intelligence-guidelines-for-non-lawyers.pdf (дата обращения: 06.09.2024).

261. Queudot M., Charton É., Meurs M. Improving Access to Justice with Legal Chatbots // Stats. — 2020. — Vol. 3(3), C. 356-375.

262. Robot Gives Guidance in Beijing Court // ChinaDaily. — URL: https://www.chinadaily.com.cn/china/2017-10/13/content_33188642.htm (дата обращения: 06.09.2024).

263. Chien Colleen V. et al. How Generative AI Can Help Address the Access to Justice Gap Through the Courts // Loyola of Los Angeles Law Review. — Forthcoming, 2024. — URL: https://ssrn.com/abstract=4683309 (дата обращения: 06.09.2024).

264. United States District Court. Southern District of New York. Case 1:22-cv-01461-PKC // CourtListener. — URL: https://storage.courtlistener.com/recap/gov.uscourts.nysd.575368/gov.uscourts.nysd.575368.54.0_6.pdf (дата обращения: 10.09.2024).

265. American Bar Association. Standing Committee on Ethics and Professional Responsibility. Formal Opinion 512. "Generative Artificial Intelligence Tools" // American Bar Association. — URL: https://www.americanbar.org/content/dam/aba/administrative/professional_responsibility/ethics-opinions/aba-formal-opinion-512.pdf (дата обращения: 12.09.2024).

266. Юристы уличили искусственный интеллект во лжи при подготовке юридических документов // Российская газета. — URL: https://rg.ru/2023/05/31/iuristy-ulichili-iskusstvennyj-intellekt-vo-lzhi-pri-podgotovke-iuridicheskih-dokumentov.html (дата обращения: 02.11.2024).

267. Выступление председателя Совета судей Российской Федерации Момотова В.В. на пленарном заседании Совета судей РФ 5 декабря 2023 года // Совет Судей Российской Федерации. — URL: http://www.ssrf.ru/news/vystuplieniia-intierv-iu-publikatsii/52649 (дата обращения: 06.09.2024).

Chapter 08

268. Artificial intelligence in healthcare. Applications, risks, and ethical and societal impacts // EPRS. — URL: https://www.europarl.europa.eu/RegData/etudes/STUD/2022/729512/EPRS_STU(2022)729512_EN.pdf (дата обращения: 30.08.2024).

269. Pham K.T., Nabizadeh A., Selek S. Artificial Intelligence and Chatbots in Psychiatry // Psychiatr Q. — 2022. — Vol. 93. — С. 249–253.

270. Du J. S. Applications of AI in Psychotherapy: An Innovative Tool // Cambridge Science Advance. — 2024. — Vol. 2024, no. 2. — С. 1–6.

271. All Worked Up. A Report on the State of American Employees' Mental Health // Wysa. — URL: https://blogs.wysa.io/wp-content/uploads/2022/12/All-Worked-Up-Report-FINAL.pdf (дата обращения: 30.08.2024).

272. Startup Uses AI Chatbot to Provide Mental Health Counseling and Then Realizes It 'Feels Weird' // Vice. — URL: https://www.vice.com/en/article/4ax9yw/startup-uses-ai-chatbot-to-provide-mental-health-counseling-and-then-realizes-it-feels-weird (дата обращения: 30.08.2024).

273. Коммерсантъ. "Как искусственный интеллект может помочь психотерапевтам" // Коммерсантъ. — URL: https://www.kommersant.ru/doc/6774969 (дата обращения: 05.09.2024).

274. Preventing AI Misuse: Current Techniques, Centre for the Governance of AI // Centre for the Governance of AI. — URL: https://www.governance.ai/post/preventing-ai-misuse-current-techniques (дата обращения: 02.09.2024).

275. Руководство по использованию генеративного искусственного интеллекта в образовании и научных исследованиях, ЮНЕСКО // ЮНЕСКО. — URL: https://unesdoc.unesco.org/ark:/48223/pf0000389639 (дата обращения: 02.09.2024).

276. Ученые разработали защиту чат-ботов от выдачи негативных советов // Российская газета. — URL: https://rg.ru/2024/01/19/uchenye-sozdali-prostoj-metod-zashchity-chat-botov-ot-vydachi-negativnyh-sovetov.html (дата обращения: 30.08.2024).

277. OpenAI тестирует ИИ-системы по модерации контента // ТАСС. — URL: https://tass.ru/obschestvo/18518001 (дата обращения: 02.09.2024).

278. Azure AI Content Safety. Safeguard user and AI-generated text and image content // Microsoft Azure. — URL: https://azure.microsoft.com/en-us/products/ai-services/ai-content-safety (дата обращения: 02.09.2024).

279. Ethics guidelines for trustworthy AI, European Commission // European Commission. — URL: https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai (дата обращения: 02.09.2024).

280. Laestadius L. et al. Too human and not human enough: A grounded theory analysis of mental health harms from emotional dependence on the social chatbot Replika. New Media & Society, 2022. — URL: https://doi.org/10.1177/14614448221142007 (дата обращения: 02.09.2024).

281. GPT-4o System Card, OpenAI // OpenAI. — URL: https://openai.com/index/gpt-4o-system-card/ (дата обращения: 02.09.2024).

282. 'I love her and see her as a real woman.' Meet a man who 'married' an artificial intelligence hologram // CBC News. — URL: https://www.cbc.ca/documentaries/the-nature-of-things/i-love-her-and-see-her-as-a-real-woman-meet-a-man-who-married-an-artificial-intelligence-hologram-1.6253767 (дата обращения: 02.09.2024).

283. The Brussels Times. "Belgian man dies by suicide following exchanges with chatbot" // The Brussels Times. — URL: https://www.brusselstimes.com/430098/belgian-man-commits-suicide-following-exchanges-with-chatgpt (дата обращения: 05.09.2024).

284. A Conversation with OpenAI's Sam Altman and Mira Murati, WSJ // The Wall Street Journal. — URL: https://www.wsj.com/podcasts/the-journal/a-conversation-with-openais-sam-altman-and-mira-murati/7c89e85f-9d7e-4569-b67d-6a777374eada (дата обращения: 02.09.2024).

285. Чрезмерное общение с чат-ботами может негативно повлиять на социализацию, Российская газета // Российская газета. — URL: https://rg.ru/2024/06/07/chrezmernoe-obshchenie-s-chat-botami-mozhet-negativno-povliiat-na-socializaciiu.html (дата обращения: 02.09.2024).

286. VC.ru. "Четыре жизни Replika: что происходит с проектом, создающим цифровую копию человека" // VC.ru. — URL: https://vc.ru/story/64057-chetyre-zhizni-replika-chto-proishodit-s-proektom-sozdayushim-cifrovuyu-kopiyu-cheloveka (дата обращения: 27.09.2024).

287. ResearchGate. “Attachment Theory as a Framework to Understand Relationships with Social Chatbots: A Case Study of Replika” // ResearchGate. — URL: https://www.researchgate.net/publication/357665581_Attachment_Theory_as_a_Framework_to_Understand_Relationships_with_Social_Chatbots_A_Case_Study_of_Replika (дата обращения: 27.09.2024).

288. Лаптев В.А. Понятие искусственного интеллекта и юридическая ответственность за его работу // Право. Журнал Высшей школы экономики. — 2019. — № 2. — URL: https://cyberleninka.ru/article/n/ponyatie-iskusstvennogo-intellekta-i-yuridicheskaya-otvetstvennost-za-ego-rabotu (дата обращения: 06.09.2024).

289. Kureha M. On the moral permissibility of robot apologies // AI & Society. — 2023. — URL: https://doi.org/10.1007/s00146-023-01782-2 (дата обращения: 03.09.2024).

290. Criminalising Offensive Speech Made by AI Chatbots in Singapore // Tech for Good Institute. — URL: https://techforgoodinstitute.org/blog/expert-opinion/criminalising-offensive-speech-made-by-ai-chatbots-in-singapore/ (дата обращения: 03.09.2024).

291. Чат-бот от Microsoft за сутки возненавидел человечество и стал нацистом // LENTA.RU. — URL: https://lenta.ru/news/2016/03/24/neonazi/ (дата обращения: 03.09.2024).

292. Валерий Зорькин высказался против наделения искусственного интеллекта правосубъектностью // Адвокатская Газета. — URL: https://www.advgazeta.ru/novosti/valeriy-zorkin-vyskazalsya-protiv-nadeleniya-iskusstvennogo-intellekta-pravosubektnostyu/ (дата обращения: 03.09.2024).

293. iConText Group. Искусственный интеллект в рекрутинге: как технологии меняют подход к поиску и найму сотрудников // Блог iConText Group. — 2024. — URL: https://blog.icontextgroup.ru/articles/iskusstvennyj-intellekt-v-rekrutinge (дата обращения: 06.09.2024).

294. OECD. Artificial Intelligence and Labour Market Matching, OECD Working Paper // OECD. — 2023. — URL: https://one.oecd.org/document/DELSA/ELSA/WD/SEM(2023)2/en/pdf (дата обращения: 04.09.2024).

295. Ekleft. Использование AI в рекрутинге: как можно улучшить и оптимизировать процесс найма // Ekleft. — 2024. — URL: https://ekleft.ru/mediatsentr/AI_HR/ (дата обращения: 06.09.2024).

296. Яков и Партнеры, Яндекс. Искусственный интеллект в России — 2023: тренды и перспективы // Яков и Партнеры. — 2023. — URL: https://yakov.partners/upload/iblock/c5e/c8t1wrkdne5y9a4nqlicderalwny7xh4/20231218_AI_future.pdf (дата обращения: 06.09.2024).

297. Искусственный интеллект Российской Федерации. Исследование HRlink: 71% HR-специалистов позитивно относится к ИИ // Искусственный интеллект Российской Федерации. — 2024. — URL: https://ai.gov.ru/mediacenter/issledovanie-hrlink-71-hr-spetsialistov-pozitivno-otnositsya-k-ii/?pageStart=150 (дата обращения: 06.09.2024).

298. Искусственный интеллект будет нанимать госслужащих: как изменится российский рынок труда? // Национальный портал в сфере ИИ. — 2024. — URL: https://ai.gov.ru/mediacenter/iskusstvennyy-intellekt-budet-nanimat-gossluzhashchikh-kak-izmenitsya-rossiyskiy-rynok-truda-/?sp (дата обращения: 04.09.2024).

299. Профессия — рекрутер: ИИ оценит личные качества претендентов на вакансию // Известия. — 2024. — URL: https://iz.ru/1729533/denis-gritcenko/professiia-rekruter-ii-otcenit-lichnye-kachestva-pretendentov-na-vakansiiu?ysclid=m0no7qtrho759589246 (дата обращения: 04.09.2024).

300. Известия. Спортивный ИИнтерес: как нейросети помогают атлетам // Известия. — 2024. — URL: https://iz.ru/1563282/alena-svetunkova/sportivnyi-iinteres-kak-neiroseti-pomogaiut-atletam (дата обращения: 09.09.2024).

301. Касиси Джоэл. Применение искусственного интеллекта в спорте // Cyberleninka. — 2024. — URL: https://cyberleninka.ru/article/n/primenenie-iskusstvennogo-intellekta-v-sporte (дата обращения: 09.09.2024).

302. Mordor Intelligence. ИИ на спортивном рынке // Mordor Intelligence. — 2024. — URL: https://www.mordorintelligence.com/ru/industry-reports/artificial-intelligence-market-in-sports (дата обращения: 09.09.2024).

303. Statista. Global sports analytics market revenue 2022-2031 // Statista. — 2024. — URL: https://www.statista.com/statistics/1185536/sports-analytics-market-size/ (дата обращения: 09.09.2024).

304. ATP Tour. Electronic Line Calling Live To Be Adopted Across The ATP Tour // ATP Tour. — 2024. — URL: https://www.atptour.com/en/news/electronic-line-calling-release-april-2023 (дата обращения: 09.09.2024).

305. Wired. DeepMind’s superhuman AI is rewriting how we play chess // Wired. — 2024. — URL: https://www.wired.com/story/deepmind-ai-chess/ (дата обращения: 09.09.2024).

Notes

Notes

Editor-in-chief I. Pozina
Project manager E. Kuznetsova
Translated by P. R. Friar, E. Kuznetsova
Layout design and cover A. Kuzmina
Layout A. Nikolaeva, O. Shchuklin

Signed for printing 29.11.2024
Format 60x90/8. SB Sans Text Headset
Coated paper. Offset printing
Print run 500 copies