Discussions on Artificial Intelligence Ethics: Development Tracks by Key Groups of Actors

приоритет2030^
лидерами становятся

MGIMO
UNIVERSITY

CONTENTS

Authors:

Anna Abramova, PhD, Director MGIMO Center for AI, Head of the Department of Digital Economy and Artificial Intelligence of the ADV group at MGIMO-University

Elena Milyaeva, specialist MGIMO Centre for AI

Maria Panova, junior researcher MGIMO Centre for AI

Anastasia Ryzhkova, PhD, researcher MGIMO Center for AI, senior consultant T1 Consulting Group

Yulia Tserekh, junior researcher MGIMO Centre for AI

Cover photo : canva.com

приоритет2030^
лидерами становятся

MGIMO
UNIVERSITY

Moscow, 2021

INTRODUCTION

Artificial intelligence systems have become the most significant technological breakthrough of the early 21st century. This technology has become the backbone of digitization and the kernel uniting all other technologies, making it possible to develop and expand the scale of the digital transformation. The widely discussed "metaverse" concept was made possible precisely through the massive introduction of various AI technologies into the manufacturing and service industries, and into people's everyday lives.

In its breakneck development and commercialization, AI is far outstripping society's capabilities to adapt the existing systems of managing processes at all levels to the new technological realities. Moreover, this technology raises the question of whether it is even possible to adapt traditional approaches to AI, and presents humankind with new challenges in the humanities and technical sciences.

Ethical issues in AI have become the leitmotif of interdisciplinary discussions of the last decade. The years 2020–2021 demonstrated that it is precisely this element of soft regulation that could shape new doctrines in management and industry development. All the main groups of actors have formed their positions on the matter, which is also reflected in this paper.

This paper attempts to outline the principal vectors in both academic and applied thought on AI ethics and the stances of the main groups of actors. We outline the already-formed academic doctrines and present the main documents that have been developed by relevant international organizations and serve as cornerstones for constructing new analytical approaches to the regulation of artificial intelligence.

We aim to reflect the main areas in the emerging interdisciplinary discussion of ethics in AI. We are deeply grateful to the leaders and coordinators of the Priority 2030 project for making it possible to conduct our research and publish our findings, which could become the first publication in a series of research papers produced at MGIMO University focused on analysing the international aspects of the development of the AI industry.

Track One: Doctrines and Their Shaping Vectors

This section describes the range of key areas where discussions are emerging on ethics in AI, lists the main academic schools working in the area, and identifies international organizations whose mandate prompts them to consider these technologies as central. In addition, it reflects the stances of businesses and non-profits that work on issues in the development and use of AI-based technologies and put their positions forward for extensive public discussions, something that appears to be of particular interest for the readers of this paper. This section also attempts to summarize the experience of leading states in developing national strategies that take ethical aspects into account. In our subsequent work, we will attempt to present as complete a picture as possible of ethics-focused AI discussions throughout the world.

THE RANGE OF KEY AREAS IN DISCUSSION DEVELOPMENT IN LEADING ACADEMIC SCHOOLS ON ETHICS IN AI

There are quite a few academic schools throughout the world that are, to some extent, involved in the matters of developing, implementing, and subsequently operating AI-based software and hardware. All the universities at the top of the QS ranking conduct AI research.

Below, we attempt to present the schools we believe to be the most impactful in the area we have selected for our paper, the schools that successfully implement an *interdisciplinary approach* that is currently largely based on studying the ethical aspects of developing and implementing AI technologies. We should immediately qualify that this list may and shall be expanded in our future work. At this stage, it is an initial attempt to introduce the lay reader to the most well-known and influential institutions. The common link in outlining the range of research institutions is, among other things, the active work of their experts in international organizations and their participation in shaping the global discussion agenda.

The Alan Turing Institute is a British national data science institute established in 2015 by five founder universities (Cambridge, Edinburgh, Oxford, Warwick, and UCL) and the Engineering and Physical Sciences Research Council.

The Institute was deliberately named after the famous British mathematician Alan Turing, who played an important part in developing informatics as a science in the United Kingdom. It is no secret that during World War II, Alan Turing led the so-called Hut 8 that worked on breaking Germany's naval codes. Turing led the team that developed both practical code-breaking methods, as well as an entire theoretical framework used to create the Bombe device that helped crack Germany's Enigma encryption machine.[1]

As part of their work on AI, the National Institute published a paper entitled "Understanding Artificial Intelligence Ethics and Safety: A Guide for the Responsible Design and Implementation of AI Systems in the Public Sector" in 2019.[2] The Guide presents its

[1] Newman, M. H. A. (1955). "Alan Mathison Turing. 1912–1954." Biographical Memoirs of Fellows of the Royal Society. 1: 253–263. doi:10.1098/rsbm.1955.0019.

[2] Leslie, D. (2019). Understanding Artificial Intelligence Ethics and Safety: A Guide for the Responsible Design and Implementation of AI Systems in the Public Sector. The Alan Turing Institute. https://doi.org/10.5281/zenodo.3240529.

readers with such quick victories in the study under research as fairness, accountability, sustainability and transparency. The paper's author believes that these areas are of key importance for developing and implementing AT-based systems.

Below, we list the key concepts from "Understanding Artificial Intelligence Ethics and Safety: A Guide for the Responsible Design and Implementation of AI Systems in the Public Sector" developed by the Alan Turing Institute.

1) *Fairness* is an approach proposed by the Institute meaning that using AI products should entail no discrimination or bias and subsequent harm to a person or society. Fairness should apply to all stages of a project's life cycle, and the essence of *fairness* is to be used for all elements involved in the project, that is, there should be "data fairness," "design fairness," "outcome fairness," "implementation fairness," etc.

2) *Accountability* is the second approach proposed by the UK's National Institute, and it means that all AI systems should be designed in such a way as to make it easy to apply end-to-end answerability and auditability of outcome. The Alan Turing Institute proposes subdividing accountability into two key types: anticipatory and remedial. "Anticipatory accountability" is applicable to the "design and development" stages of a software development project. The second type, "remedial accountability," is to be used at the immediate implementation stage.

3) *Sustainability* is the third principle of developing AI technologies proposed by the Alan Turing Institute, and it means that both the developers and users of AI-based systems should know and remember that the direct use of AI-based software directly transforms the behavior, opinions and stances of both individuals and society as a whole, as it erases the boundaries between the real and digital worlds. This is why the paper's author believes that it is crucial for both developers of AI-based systems and their immediate users to remain "sensitive" to changes in the real world. It is important to understand that both the survival and robustness of AI-based systems depend on compliance with operational goals that are, in turn, connected with security that includes such performance indicators as "accuracy," "reliability," "security" and "robustness."

4) *Transparency* is the final concept in the Alan Turing Institute's paper "Understanding Artificial Intelligence Ethics and Safety: A Guide for the Responsible Design and Implementation of AI Systems in the Public Sector." It means the need to justify the ethical permissibility of using AI-based software. This principle also entails preventing discrimination, meaning that there is "public trust" both in the outcome and in the technologies that underlie this outcome. The author additionally emphasizes that all parties concerned should have access to both explanations and clarifications of operating principles and outcomes achieved.

This paper has laid the foundation for subsequent discussions that were continued within projects and papers such as "Ethics of Machine Learning in Children's Social Care,"[3] which focuses on aspects of using machine-learning algorithms to offer more narrowly targeted social care to children, or the "AI for Multiple Long-Term Conditions Programme intended to design methodological approaches to using AI via safe and comprehensible infrastructure, using prepared data, training qualified personnel, conveying the practice of using AI to the public at large, and developing sustainable development principles.

Additionally, it is important to know that ethical matters also pertain to data usage, data preparation, data labeling, and the subsequent use of data, which directly influences the development of the AI industry.

The Atomium–European Institute for Science, Media and Democracy conducts systemic research on AI. One of the Institute's most important and talked-about studies is the "AI4People's Ethical Framework for a Good AI Society: Opportunities, Risks, Principles, and Recommendations," which submitted for a broad public discussion in November 2018 at the multilateral AI4People forum. The paper

[3] https://www.turing.ac.uk/research/research-projects/ethics-machine-learning-childrens-social-care.

analysed the existing ethical principles for using AI (a total of 47 principles from six documents, as well as other principles developed by outside organizations and enshrined in the relevant reports, declarations, and statutes) and suggested synthesizing them.

We believe it important to note that four out of five principles proposed by AI4People are used in "bioethics" (Bioethics is an interdisciplinary research area that applies to the moral aspect of human activities in medicine and biology that emerged in the mid-20th century at the conjunction of philosophical disciplines, law, and natural sciences[4]):

- *Beneficence:* promoting well-being, preserving dignity, sustaining the planet
- *Non-maleficence:* privacy, security and "capability caution"
- *Autonomy:* the power to decide
- *Justice:* promoting prosperity and preserving solidarity [5]

The authors of "AI4People's Ethical Framework for a Good AI Society: Opportunities, Risks, Principles, and Recommendations" also proposed a fifth principle, *"explicability,"* which subsumes other principles related to "intelligibility" and "accountability."

It should also be noted that the paper contains 20 recommendations on such procedures as "assessment," "development," "incentivization," "support" for developing "good" AI, etc.

In 2019, the Institute analyzed issues related to building a balanced system of regulating the AI industry, with the subsequent establishment of seven sectoral committees.[6]

The Berkman Klein Center for Internet & Society at Harvard University (United States). The paper "Principled Artificial Intelligence: Mapping Consensus in Ethical and Rights-Based Approaches to Principles for AI"[7] presented in 2020 deserves a special mention among the Center's publications. The paper contains the findings derived from the Center's analysis of 36 studies on the principles of the AI development worked out in various countries reflecting the interests of different stakeholders.

The first aspect of the research is the identification of eight key themes:

- *Privacy.* This comprises eight principles: "consent," "control over the use of data," "ability to restrict processing," "right to rectification," "right to erasure," "privacy by design," "recommends data protection law," and "privacy (other/general)."
- *Accountability.* This comprises ten principles: "verifiability and replicability," "impact assessments," "environmental responsibility," "evaluation and auditing requirement," "creation of a monitoring body," "ability to appeal," "remedy for automated decision," "liability and legal responsibility," "recommends adoption of new regulations," and "accountability per se."
- *Safety and security.* This comprises four principles: "safety," "security," "security by design," and "predictability."
- *Transparency and explainability.* This comprises eight principles: "transparency," "explainability," "open source data and algorithms," "open government procurement," "right to information," "notification when AI makes a decision about an individual," "notification when interacting with AI," and "regular reporting."
- *Fairness and non-discrimination.* This comprises six principles: "non-discrimination and the prevention of bias," "representative and high quality data," "fairness," "equality," "inclusiveness in impact," and "inclusiveness in design."

[4] Bioethics and Biotechnologies: Limits for Improving the Human Being. A Festschrift for Pavel Tishchenko on His 70th Anniversary. E. G. Grebenshchikova, B. G.Yudin, eds. Moscow: Moscow Humanities University Press, 2017. 240 p.

[5] https://www.eismd.eu/wp-content/uploads/2019/02/Ethical-Framework-for-a-Good-AI-Society.pdf.

[6] AI4People's Ethical Framework for a Good AI Society: Opportunities, Risks, Principles, and Recommendations. 2019. https://www.eismd.eu/wp-content/uploads/2019/03/AI4People%E2%80%99s-Ethical-Framework-for-a-Good-AI-Society.pdf.

[7] Fjeld, Jessica, Nele Achten, Hannah Hilligoss, Adam Nagy, and Madhulika Srikumar, "Principled Artificial Intelligence: Mapping Consensus in Ethical and Rights-Based Approaches to Principles for AI". Berkman Klein Center Research Publication No. 2020-1 (January 15, 2020). Available at SSRN: https://ssrn.com/abstract=3518482.

- *Human control of technology.* This comprises three principles: "human review of automated decision," "ability to opt out of automated decision," and "human control of technology (other/general)."
- *Professional responsibility.* This comprises five principles: "accuracy," "responsible design," "consideration of long term effects," "multi-stakeholder collaboration," and "scientific integrity."
- *Promotion of human values.* This comprises three principles: "human values and human flourishing," "access to technology," and "leveraged to benefit society."

It should be noted that consideration of the humanities-related aspects of developing AI technologies raises the crucial dilemmas of developing cutting-edge technologies: the work of a mathematically calculated algorithm appears to many to be a black box with information put in and coming out. How do we reflect in this work the ethical norms of morality, fairness, equality, responsibility, non-discrimination, which are of crucial importance today?

Therefore, we may say that over the past few years, the leading academic schools have already formulated key "values" that should be translated into the language of algorithms in the near future, and should also be expressed in the applied evaluative approaches that are relevant for technological companies of various levels.

AI ETHICS WITHIN THE NATIONAL STRATEGIES OF STATES: THE EXPERIENCE OF LEADING STATES

To open this chapter, we would like to note that 34 national AI development strategies had been adopted worldwide as of May 2021.

National AI strategies attract the attention of various researchers, including those who attempt to formulate a kind of system that would unite the strategies of different countries.

Of particular interest is a study published by the Brookings Institution,[8] which identifies four groups of "signals" that all states with approved national AI strategies give off to one degree or other:

1. *Traditional signals* are both deliberate and true. An example of a traditional signal is a Strategy mentioning the accurate amount of investment made in AI research.
2. *Inadvertent disclosure signals:* such strategies transmit true information, but do not do so deliberately. An example would be a country planning to spend a great deal of money on infrastructure, which reveals a belief that the country is deficient in that area.
3. *Opportunistic signals* are not true but are sent deliberately as a rule. For instance, the government of a country might say it intends to use AI for developing public services but actually intends to use it for warfare-based systems.
4. *Mixed signals* unintentionally transmit false information in national strategies. An example of a mixed signal in AI plans is the declaration of using anonymized public data in AI systems but failure to do so.

The concept of soft law is used with increasing frequency in discussions of approaches to building national and international management for the AI industry. This may include the development of principles, guidelines, certification and standardization, and, certainly, codification. The approach itself is neither new nor unique. Interest in it increases cyclically when the agenda comes to feature a new, rapidly developing technology with the potential to be used across a broad range of economic sectors. At such times, it becomes evident that the current regulatory framework is insufficient, while the accelerated adoption of new legislative initiatives without a proper in-depth study of the new technology risks limiting scientific and technological progress.

[8] https://www.brookings.edu/blog/techtank/2021/05/13/analyzing-artificial-intelligence-plans-in-34-countries/.

The strength of this aspect lies in the high variability of solutions with different configurations of market participants. At the same time, the main actors from all the parties concerned may be involved or, conversely, the document may be geared towards managing a narrow problem in a specific sector. Given the state of AI industry development, such capabilities make soft law a highly desirable approach.

In 2021, Carlos Ignacio Gutierrez and Gary E. Marchant of Arizona State University conducted a large-scale study to summarize all the existing soft law projects and determine patterns, including regional patterns.[9] Their research demonstrated that the proposed methodology allows for 634 projects to be selected. The study's timeframe spanned two decades, and it turned out that approximately 90% of the initiatives had been proposed in 2016–2019. This coincides with the stage of actively commercializing AI technologies and implementing them across a broad range of sectors in all countries that had identified AI as a strategic priority. The accumulated data shows that soft law is most widespread in high-income countries, predominantly the United States and European states. The study shows that nearly 55% of all initiatives are AI recommendations and strategies, followed by AI principles (nearly 25%) and standards (9.5%). The top four is rounded off by professional guidelines and codes of conduct (nearly 4%).

Below, we give examples of leading states constructing such national strategies and initiatives for developing and commercializing AI technologies that the authors believe to be most significant for developing the area under study.

It is important to know that the Russian Federation adopted its own national AI development strategy in 2019. However, given the international thrust of our Overview, we leave the consideration of Russian practices outside the scope of our research in order to study it in more detail and assess its outcomes in subsequent analytical papers.

The United States of America

The United States adopted its National Artificial Intelligence Initiative on January 1, 2021.[10]

The key objective of the National Artificial Intelligence Initiative is to "ensure continued U.S. leadership in AI R&D; lead the world in the development and use of trustworthy AI systems in public and private sectors; prepare the present and future U.S. workforce for the integration of artificial intelligence systems across all sectors of the economy and society."

The National Artificial Intelligence Initiative takes a comprehensive approach to synergy and research coordination, AI development and education across all U.S. departments and agencies, and involving academic circles, industry, non-profits, and civil society organizations in the cooperation. Work under this initiative is divided into six strategic areas: innovations, promoting safe AI, education and training, infrastructure, apps, and international cooperation.

The United Kingdom of Great Britain and Northern Ireland

The United Kingdom of Great Britain and Northern Ireland approved its National AI Strategy in September 2021.[11] It reflects the key stages of the country's ten-year plan to shaping itself as a global AI superpower.

The document's authors believe that the Strategy should have the following key areas:

- Invest and plan for the long-term needs of the AI ecosystem to continue the UK's leadership as a science and AI superpower.
- Support the transition to an AI-enabled economy, capturing the benefits of innovation in the UK, and ensuring AI benefits all sectors and regions.
- Ensure the UK gets the national and

[9] Gutierrez, Carlos Ignacio, and Gary E., Marchant, "A Global Perspective of Soft Law Programs for the Governance of Artificial Intelligence" (May 27, 2021). Available at SSRN: https://ssrn.com/abstract=3855171 or http://dx.doi.org/10.2139/ssrn.3855171.

[10] https://www.ai.gov.

[11] https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/1020402/National_AI_Strategy_-_PDF_version.pdf.

international governance of AI technologies right to encourage innovation, investment, and protect the public and our fundamental values.

Therefore, a characteristic feature of the Anglo-Saxon approach is moving away from including individual ethical provisions in AI in their strategic framework documents. However, the United States, United Kingdom and Australia have already developed and are applying sectoral documents, including documents on military applications of AI.[12] It should be noted, however, that large research centers with a strong reputation at all levels of state governance make a major contribution to promoting the ethical agenda.

The People's Republic of China

China began to actively promote AI in 2017, when the State Council published its "New Generation AI Development Plan,"[13] which envisions "China's transformation into a leading AI power by 2030."[14]

Subsequently, in 2018–2020, China published:[15]

- The White Paper on AI Standardization prepared by the Standardization Administration of China and China Electronics Standardization Institute;
- Baidu, a Chinese web services company that includes a search engine of the same name, published four AI ethics principles that include: security and manageability, equal access to technology and opportunities, AI for good and the growth and development of the people, and more freedom and opportunities for humanity;
- The ARCC principles, claiming that AI should be accessible, safe, understandable, and manageable;
- The White Paper on AI Security prepared by the China Academy of Information and Communications Technology;
- Beijing AI principles to be followed in researching the first steps in codifying and developing AI (AI serves the good of the humanity and the environment, serves human values, is used responsibly, ethically, inclusively, openly, and all the risks are controlled), using AI (its use should be wise and proper, involve informed consent, education and training), and AI governance (it should involve harmony and cooperation, adaptation and moderation, subdivision and implementation, long-term planning in AI, and optimizing employment).[16] These principles call for "the construction of a human community with a shared future, and the realization of beneficial AI for humankind and nature";[17]
- The Ministry of Science and Technology published its "Governance Principles for the New Generation of AI," which include "Harmony and Friendliness," "Fairness and Justice," "Inclusiveness and Sharing," "Respect for Privacy," "Security and Controllability," "Shared Responsibility," "Open Cooperation," and "Agile Governance";
- The National AI Standardization Group's report on analyzing ethical risks in AI;
- The Artificial Intelligence Industry Alliance's "joint pledge" on self-discipline in the AI industry;
- The main principles of the Megvii tech company's artificial intelligence practice;
- The AI Security Standardization White Paper prepared by the National Information

[12] Will Douglas Heaven, "The Department of Defense is Issuing AI Ethics Guidelines for Tech Contractors." https://www.technologyreview.com/2021/11/16/1040190/department-of-defense-government-ai-ethics-military-project-maven/; "Understanding Artificial Intelligence Ethics and Safety." https://www.turing.ac.uk/research/publications/understanding-artificial-intelligence-ethics-and-safety; "Artificial Intelligence. An Accountability Framework for Federal Agencies and Other Entities." GAO 2021. https://www.gao.gov/assets/gao-21-519sp.pdf.

[13] Asia's AI Agenda: The Ethics of AI. MIT Technology Review Insights, 2019. https://www.technologyreview.com/2019/07/11/134229/asias-ai-agenda-the-ethics-of-ai/.

[14] AI Policy – China. Future of Life Institute. 2018. https://futureoflife.org/ai-policy-china/.

[15] Rebecca Arcesati, "Lofty Principles, Conflicting Incentives: AI Ethics Governance in China," Mercator Institute for China Studies (2021). https://merics.org/en/report/lofty-principles-conflicting-incentives-ai-ethics-and-governance-china.

[16] The Beijing Artificial Intelligence Principles. Wired – The Latest in Technology, Science, Culture and Business. 2019. https://www.wired.com/beyond-the-beyond/2019/06/beijing-artificial-intelligence-principles/.

[17] AI Policy – China. Future of life institute. 2018. https://futureoflife.org/ai-policy-china/.

Security Standardization Technical Committee of China (TC260);

- The first civil lawsuit over using face recognition technology;
- Private life should be publicly inviolable, which means opposing the use of ZAO deepfake app (where neural networks can be used to substitute a person's face with another in any film or video);
- The White Paper on AI Governance prepared by the China Academy of Information and Communications Technology and China's Artificial Intelligence Industry Alliance;
- The Beijing Principle of Artificial Intelligence for Children designed by the Beijing AI Academy. The Beijing Consensus of Artificial Intelligence for children represents the first guidelines published in China on developing AI for children. Particular attention is paid to children-centred values, child protection, adopting commitments and introducing multilateral governance. These topics cover 19 detailed principles, including dignity, fairness, putting children first, protecting privacy, taken the will of children into account, etc.[18]

China's experience thus demonstrates that it is possible to combine framework initiatives and a sectoral approach with specific ethical approaches in sufficiently narrow areas.

The Commonwealth of Australia

The authors of the Commonwealth of Australia's National AI Development Strategy believe that it should serve as the foundation for positioning the country as "a global leader in developing and adopting trusted, secure and responsible AI."

Crucially, the AI Action Plan is integrated into the Australian government's National Digital Economy Strategy.

The national plan is based on four "focuses":

1) *Transforming Australian businesses* – plans involve forming and implementing public support measures for businesses to develop and implement AI technologies with a view to creating jobs, improving productivity and giving them a competitive advantage.
2) *Creating an environment to grow and attract the world's best AI talent* – government support for businesses to ensure access to world-class talent and experience.
3) Using cutting-edge AI technologies to solve *Australia's national challenges* – support for Australia's world-leading capabilities in AI research to solve national challenges and ensuring that all Australians may benefit from the advantages afforded by AI.
4) *Making Australia a global leader in responsible and inclusive AI* – support for AI inclusivity and developing technologies that reflect Australian values.

Therefore, the framework document does not directly introduce ethical provisions, but it does involve certain sectoral initiatives.[19] It also introduces framework approaches to the application of ethical norms in developing and using AI technologies.[20] It is interesting to note that the principles were tested in large Australian companies. Some companies already had their own practices and methodologies, while others decided to implement only a part of what had been proposed. A TNC also joined the process at the national level, namely Microsoft, which already had its own AI ethics practices.

INTERNATIONAL ORGANIZATIONS AND THEIR STANCES ON ETHICAL ISSUES IN AI

The exponential growth of digital technologies and artificial technologies accelerates international cooperation in developing ethical approaches. It also emphasizes the importance

[18] Younas, Ammar, "The Beijing Consensus of Artificial Intelligence for Children: An Effort to Prevent Juvenile Delinquency" (September 19, 2020). Available at SSRN: https://ssrn.com/abstract=3695631 or http://dx.doi.org/10.2139/ssrn.3695631.

[19] Ethical Principles for AI in Medicine. The Royal Australian and New Zealand College of Radiologists. August 2019.). // URL: https://www.ranzcr.com/college/document-library/ethical-principles-for-ai-in-medicine.

[20] Australia's Artificial Intelligence Ethics Framework. https://www.industry.gov.au/data-and-publications/australias-artificial-intelligence-ethics-framework.

of a global approach and transitioning from the theoretical development of principles to their practical application. In particular, that was the opinion stated by the attendees of the International Conference on AI Ethics organized by the Council of Europe and Hungary (namely, representatives of the OECD, UNESCO, the European Commission and the Council of Europe's Ad Hoc Committee on Artificial Intelligence), which took place on October 26, 2021.[21]

International organizations such as UNESCO, the OECD, the Council of Europe (in particular its Ad Hoc Committee on Artificial Intelligence), UNCTAD, the European Commission, and the WHO make a significant contribution to developing the discussions on ethics in AI.

Organisation for Economic Co-operation and Development (OECD)

For several decades, the OECD has been conducting analytical research into information economy and its key sectors. Currently, it is assessing the problems and prospects of digital economy development from various angles. The AI industry has become one of the central areas of this research. Analysing the AI business, OECD experts used a systemic approach that allowed them to cover a large range of problems including ethical aspects. The Recommendation of the Council on Artificial Intelligence was adopted on May 22, 2019, [22] the first intergovernmental document on the standardization of approaches. The document itself states that "The Recommendation aims to foster innovation and trust in AI by promoting the responsible stewardship of trustworthy AI while ensuring respect for human rights and democratic values. Complementing existing OECD standards in areas such as privacy, digital security risk management, and responsible business conduct, the Recommendation focuses on AI-specific issues and sets a standard that is implementable and sufficiently flexible to stand the test of time in this rapidly evolving field. In June 2019, at the Osaka Summit, G20 Leaders welcomed G20 AI Principles, drawn from the OECD Recommendation."[23]

The OECD Recommendation presents five *principles for the responsible stewardship of trustworthy AI:*

- inclusive growth, sustainable development and well-being;
- human-centred values and fairness;
- transparency and explainability;
- robustness, security and safety;
- accountability.

The OECD AI Policy Observatory (OECD. AI)[24] established in 2020 is based on the OECD Recommendation on Artificial Intelligence.

The Observatory's website says that "OECD. AI combines resources from across the OECD, its partners and all stakeholder groups. OECD. AI facilitates dialogue between stakeholders while providing multidisciplinary, evidence-based policy analysis in the areas where AI has the most impact."[25]

Additionally, the OECD Directorate for Science, Technology and Innovation (STI) regularly publishes research on digital technologies (reports, articles, etc.).[26] These works reflect, with varying degrees of detail, the key areas of discussion that different groups of actors conducted on issues in building a balanced governance system for the digital transformation and developing new technology markets. Work on classifying AI systems is currently under way. Risk analysis is considered as the next possible step.

The United Nations Educational, Scientific and Cultural Organization (UNESCO)

In November 2021, UNESCO adopted its Recommendation on the Ethics of Artificial Intelligence, which had been in development since 2019.[27] The document was signed by 193 UNESCO member states and adopted pursuant to extended expert discussions.

[21] "Current and Future Challenges of Coordinated Policies on AI Regulation": International Conference. https://www.coe.int/en/web/artificial-intelligence/-/-current-and-future-challenges-of-coordinated-policies-on-ai-regulation-international-conference.

[22] https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0449.

[23] https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0449#backgroundInformation.

[24] https://oecd.ai/en/.

[25] https://oecd.ai/en/about.

[26] https://www.oecd-ilibrary.org/science-and-technology/oecd-digital-economy-papers_20716826.

[27] Recommendation on the Ethics of Artificial Intelligence https://unesdoc.unesco.org/ark:/48223/pf0000379920.page=14.

The Recommendation states that its purpose is to lay the groundwork for using AI for the benefit of people and nature and to stimulate the use of AI solely for peaceful purposes. The document lists the following objectives:

- to provide a universal framework of values, principles and actions to guide States in the formulation of their legislation, policies or other instruments regarding AI, consistent with international law;
- to guide the actions of individuals, groups, communities, institutions and private sector companies to ensure the embedding of ethics in all stages of the AI system life cycle;
- to protect, promote and respect human rights and fundamental freedoms, human dignity and equality, including gender equality; to safeguard the interests of present and future generations; to preserve the environment, biodiversity and ecosystems; and to respect cultural diversity in all stages of the AI system life cycle;
- to foster multi-stakeholder, multidisciplinary and pluralistic dialogue and consensus building about ethical issues relating to AI systems;
- to promote equitable access to developments and knowledge in the field of AI and the sharing of benefits.

The Recommendation also covers value paradigms, technological operating principles and AI systems. The document lists the following priorities: ethical impact assessment, ethical management and governance, data policies, development and international cooperation, the environment and ecosystems, gender equality, culture, education and research, communications and information, economy and labour market, health and social welfare.

As Ms. Gabriela Ramos, Assistant Director-General for Social and Human Sciences, says, "The Recommendation on the Ethics of Artificial Intelligence will be a blueprint for global consensus on the 'what,' as well as the 'how' of ethical regulation of this game-changing technology. UNESCO stands ready to assist governments and other stakeholders in developing their capacities to address the challenges, including through the ethical impact assessment."[28] The issues of monitoring and assessment prompted the largest number of expert comments owing to the proposed instrument being unclear and non-transparent.

The Council of Europe Ad Hoc Committee of Artificial Intelligence

The Ad Hoc Committee on Artificial Intelligence (CAHAI) contributes significantly to the development of AI ethical principles.

Convention 108[29] of the Council of Europe supplemented and amended in the 2018 Protocol (Convention 108+[30]) establishes global standards in human rights to privacy and data protection regardless of the level of technological development. In particular, the document implies processing special categories of data (confidential data) only when relevant circumstances are enshrined in the law that supplements terms stipulated in the Convention, and grants every person the right to know that their personal data are being processed and allows them to make changes to these data or erase them completely if processing such data contradicts the provisions of the Convention. The Protocol amending the Convention added new principles such as processing being transparent (Article 8), proportionate (Article 5), accountability (Article 10), such as examining likely impact (Article 10) and respect for privacy by design (Article 10). The Council of Europe notes that these new rights have a special significance in regard to people's profiling and automated decision-making.[31]

In 2020, the Ad Hoc Committee for Artificial Intelligence prepared a new Progress Report. [32] Current AI ethical guidelines throughout the world converge on certain common points, but

[28] https://en.unesco.org/news/ai-ethics-another-step-closer-adoption-unescos-recommendation-0
[29] https://rm.coe.int/1680078b37
[30] https://www.coe.int/en/web/conventions/full-list?module=treaty-detail&treatynum=223
[31] https://search.coe.int/cm/Pages/result_details.aspx?ObiectID=09000016809fa65b
[32] https://search.coe.int/cm/Pages/result_details.aspx?ObiectID=09000016809fa65b

the countries diverge sharply on the details of what should actually be done. In particular, the document says that as regards transparency (the principle that is most frequently defined), it was unclear whether it should be achieved by publishing the source code, giving access to algorithm learning principles or by auditing them (with account for personal information protection laws) or through some other means. Solving the problem of applying these principles in practice and considering potential correlations and trade-offs with other desirable objects was, consequently, deemed to be an important question to be handled by policy makers.

The Ad Hoc Committee on Artificial Intelligence also prepared a Feasibility Study[33] that includes "AI Ethics Guidelines: European and Global Perspectives."[34]

The research aims to map soft law norms and other ethical and legal documents developed by government agencies and NGOs throughout the world with a view to simplifying the monitoring of such legal frameworks and promptly tracking and assessing the impact AI has on ethical principles, human rights, the rule of law, and democracy. The document contains an overview of 116 papers from various states.

The United Nations Conference on Trade and Development (UNCTAD)

UNCTAD has published regular overviews for a number of years now. Initially, these overviews reflected the formation of the information society in all groups of countries. Later, as the concept of digital economy developed, they published biennial overviews of digital economy development. In 2021, the organization presented its Digital Economy Report subtitled "Cross-Border Data Flows and Development: For Whom the Data Flow."[35] In the preface to the report, UN Secretary-General Antonio Guterres says that "The Report calls for innovative approaches to governing data and data flows to ensure more equitable distribution of the gains from data flows while addressing risks and concerns. A holistic global policy approach has to reflect the multiple and interlinked dimensions of data and balance different interests and needs in a way that supports inclusive and sustainable development with the full involvement of countries trailing behind in digital readiness."[36] Problems in AI technology development are viewed through the lens of issues in building a data transfer management system where ethics is a key element.

The Report notes the need to develop national strategies for data and their cross-border flows, strategies that "can help reap economic development gains, while at the same time respecting human rights and various security concerns. Third, capacity-building activities may be needed to raise awareness of data-related issues and their development implications."[37] UNCTAD also calls for global management of data and their cross-border flows and points out the importance of international technical coordination. UNCTAD emphasizes that "there is a race for leadership in digital technologies developments, as it is thought that controlling the data and related technologies, particularly artificial intelligence, will secure economic and strategic power."[38] The Report also stresses the need to discuss ethical aspects related to data analysis and AI development.

The European Commission

This organization systemically works on building analytical approaches to regulating AI systems. In 2019, the European Commission published the mutually complementary Ethics Guidelines for Trustworthy AI and Ethics Recommendation for Trustworthy AI in Politics and Investment.

The organization formulated seven AI requirements:[39]

[33] https://rm.coe.int/cahai-202канада0-23-final-eng-feasibility-study-/1680a0c6da

[34] https://rm.coe.int/cahai-2020-07-fin-en-report-ienca-vayena/16809eccac

[35] https://unctad.org/system/files/official-document/der2021_overview_ru.pdf.

[36] Ibid.

[37] Ibid.

[38] Ibid.

[39] https://ec.europa.eu/commission/presscorner/detail/en/IP_19_1893.

- Human agency and oversight
- Robustness and safety
- Privacy and data governance
- Transparency
- Diversity, non-discrimination and fairness
- Societal and environmental well-being
- Accountability

"The ethical dimension of AI is not a luxury feature or an add-on. It is only with trust that our society can fully benefit from technologies," stated the European Commission's Vice-President for the Digital Single Market Andrus Ansip.[40]

The World Health Organization (WHO)

Medicine made the short list of sectors where AI technologies quickly found wide commercial application. Moreover, it is here where the ethical aspects of using AI technologies are particularly prominent. In 2021, the World Health Organization presented its "Ethics & Governance of Artificial Intelligence for Health,"[41] which identified six principles to be used as guidelines in developing and implementing AI tools in healthcare:[42]

- Protecting human autonomy
- Ensuing transparency, explainability, and intelligibility
- Promoting human well-being and safety and the public interest
- Fostering responsibility and accountability
- Ensuring inclusiveness and equity
- Promoting AI that is responsive and sustainable

These principles were the result of a consensus discussion among experts in medicine, law, ethics, and digital technologies. The report also contains recommendations on improving AI governance systems in the public and private sectors.

NON-PROFITS AND THEIR CONTRIBUTION TO DEVELOPING THE DISCUSSIONS ON ETHICS IN AI

This section presents an analysis of initiatives in AI ethics proposed by the non-profit sector. We have considered the key proposals put forward by laboratories and non-profits researching AI that are most frequently cited in open sources. Given that, until recently, issues in ethical use of AI were the most highly debated, we emphasized this particular issue in our coverage of non-profits.

Below, we provide a short overview of the activities of the main non-profits that had a particular impact at the leading discussion venues covering the problems of AI development.

The AI Ethics Lab focuses on consulting and research. It "aims to detect and address ethics risks and opportunities in building and using AI systems to enhance technology development."[43] Since 2017, it has been implementing its bespoke PiE (puzzle-solving in ethics) model, which it uses to apply such ethical instruments and solutions as an AI ethics roadmap, AI ethics strategy, AI ethics analysis, and education in AI ethics.

The Association for Computing Machinery brings together educators, researchers, and computing machinery professionals. The Association has an Ethics and Professional Conduct Code, and in 2017, it issued a Statement on Algorithmic Transparency and Accountability. This statement concerns seven principles that agree with the Code and are intended "to support the benefits of algorithmic decision-making" in problem-solving to "minimize potential harms while realizing the benefits of algorithmic decision-making."[44] These principles include:

[40] Ibid.

[41] https://www.who.int/publications/i/item/9789240029200.

[42] https://www.who.int/ru/news/item/28-06-2021-who-issues-first-global-report-on-ai-in-health-and-six-guiding-principles-for-its-design-and-use.

[43] https://aiethicslab.com/about/.

[44] https://www.acm.org/binaries/content/assets/public-policy/2017_usacm_statement_algorithms.pdf.

- Awareness
- Access and redress
- Accountability
- Explanation
- Data provenance
- Auditability
- Validation and testing

The Future of life Institute is a Boston-based charity working to reduce the risks associated with AI, nuclear weapons and biotechnologies. The 2017 Asilomar Conference developed a list of 23 principles for working with AI that range from research strategies to rights to data, and future problems.[45] A separate "Ethics and Values" section covers the following principles: safety, failure transparency, judicial transparency, responsibility, value alignment, human values, personal privacy, liberty and privacy, shared benefit, shared prosperity, human control, non-subversion, and AI arms race.[46]

The Institute for Ethical AI & Machine Learning is a British analytical centre working on industry standards for data management and machine learning. The centre developed eight principles that "provide a practical framework to support technologists when designing, developing or maintaining systems that learn from data."[47] These machine learning principles include:

- Human augmentation
- Bias evaluation
- Explainability by justification
- Reproducible operations
- Displacement strategy
- Practical accuracy
- Trust by privacy
- Data risk awareness

The Linux Foundation supports open technologies projects in the development of world-class open-source software, communities and companies. In 2021, the Working Group on Trustworthy AI Principles Committee announced LF AI & Data principles. These principles form the acronym (R) REPEATS, which stands for Reproducibility, Robustness, Equitability, Privacy, Explainability, Accountability, Transparency, and Security.

The Rome Call for AI Ethics. In February 2020, the Pontifical Academy for Life (Vatican), in collaboration with Microsoft, IBM, FAO, and Italy's Ministry of Innovation Technology signed the "Call for an AI Ethics," a document developed to support an ethical approach to AI, increase the sense of responsibility among organizations, governments, institutions, and the private sector and promote a future where AI technologies are human-oriented and subordinated to people instead of replacing them.[48] The document comprises three areas: ethics, education, and rights; and includes six principles: transparency, inclusion, responsibility, impartiality, reliability, security and privacy.

AI Ethics Principles & Guidelines by Smart Dubai is a set of AI ethics tools (comprising principles and recommendations) developed in Dubai to assist the urban ecosystem in responsibly using an AI system. AI principles are:

- Ethics: AI systems should be fair, transparent, accountable and understandable;
- Security: AI systems should be safe and secure, and should serve and protect humanity.
- Humanity: AI should be beneficial to humans and aligned with human values, in both the long and short term.
- Inclusiveness: AI should benefit all people in society, be governed globally, and respect dignity and people rights.[49]

We have analysed the activities of various non-profits in the development and introduction of ethical principles into creating and implementing AI technologies. Our key takeaways are as follows:

1. Even though there is no principle that appears in all guidelines, transparency, fairness,

[45] https://futureoflife.org/2017/01/17/principled-ai-discussion-asilomar/.
[46] https://futureoflife.org/ai-principles-russian/.
[47] https://ethical.institute/principles.html.
[48] https://www.romecall.org/.
[49] https://www.digitaldubai.ae/docs/default-source/ai-principles-resources/ai-ethics.pdf?sfvrsn=d4184f8d_6.

impartiality, non-maleficence, responsibility, and privacy are mentioned in over the half of them. This may be considered evidence of ethical AI gradually converging around these principles in global policies. In particular, the dominance of calls for transparency and fairness indicates a growing moral priority that requires transparent processes throughout the entire operations of an AI cycle (from transparency in developing and designing algorithms to the transparent use of AI).

2. Principles are a valuable part of any applied ethics, as they help reduce complicated ethical issues to several central elements that can ensure widespread commitment to a shared set of values. Although these principles do indeed reflect consensus concerning the important and desirable goals in developing and using technologies, they do not provide practical recommendations for understanding new and complicated situations.

3. The question remains open: Will a unified set of ethical AI principles be adopted and applied throughout the world and will it reflect the interests and needs of all members of society?

Track Two: First Steps in Codifying Ethics in Artificial Intelligence

Even though codification in professional activities has a long history, this area in AI ethics has come into focus of discussions only in the last two to three years. It has become evident that the expanding range of AI principles has already spanned all the principal participants, but that does not help prevent the negative consequences of using AI technologies. At the same time, states are not ready to introduce harsh legal regulations. What is required is a balanced approach that will allow us to preserve a human-centric view while refraining from introducing rigid restrictions on technological progress. Codifying ethics in AI has become such an option. This soft form of regulation has great potential today in various areas and at various levels of actor involvement.

CHINA'S CODE OF ETHICS IN ARTIFICIAL INTELLIGENCE

The first comprehensive code of ethics in artificial intelligence was developed in China. On September 25, 2021, the New Generation Artificial Intelligence Strategic Advisory Committee at the Ministry of Science and Technology of the People's Republic of China published a **"New Generation Artificial Intelligence Code"** that applies to all organizations and enterprises that use AI in China.[50] This was another step in China's journey towards **global leadership** in AI development by 2030.

The Code has six sections and 25 articles that contain general provisions, R&D, governance and use norms, and concluding provisions covering the organization and implementation of the Code's norms.

China's code includes the following principal provisions:

1. The use of AI technologies should be guided by the principles of improving human well-being, promoting veracity and fairness, protecting privacy and security, ensuring governability and reliability, increasing responsibility, and improving ethics literacy (Article 3).

2. The functioning of AI technologies, including issues in AI ethics, is regulated by law: AI-related regulations (laws, standards, policies) must be complied with, and AI ethics must be integrated into the governance process (Article 6). AI with its potential, development vector and limitations (Article 5) should be carefully balanced against a human person, whose rights and freedoms should not be diminished or violated (Article 7).

3. A human person remains free in making decisions on using AI technologies (Article 3) and may at any time refuse to interact with AI and suspend the systems' operations. This means that control over AI functioning and responsibility for it always lies with a human person.

4. Market rules must be complied with when using AI (Article 14). Monopolies of any kind are not permitted since they violate competition and intellectual property rights.

5. AI products must be used in good faith (Article 18). Using them for anything other than their intended purpose (Article 20), misuse or abuse (Article 19) must be avoided. Actions involved in working with AI should not be detrimental to the rights and interests of users, society, the state, and national security.

6. Actors, such as departments, enterprises, universities, research institutions,

[50] https://www.globalgovernmentforum.com/china-unveils-ai-ethics-code/.

associations, and other agencies can develop their own AI codes using the official code as their base document.

Experts note that China strives to tighten state control over its technological sector. Recently, it adopted strict measures on recommendation engine algorithms[51] and tightened rules on user information.[52] Freedom House, an independent democracy watchdog organization, ranked China as the world's worst country in terms of the freedom of the Internet for the seventh year running.[53,54]

RUSSIA'S CODE OF ETHICS IN ARTIFICIAL INTELLIGENCE

ЭТИКА
ИСКУССТВЕННОГО
ИНТЕЛЛЕКТА:
НАЧАЛО ДОВЕРИЯ

Кодекс этики в сфере искусственного интеллекта

Figure 1: Russia's Code of Ethics in Artificial Intelligence

Russia presented its code of ethics in AI on October 26, 2021.[55] Its publication followed a year of preparations and searching for a compromise between the principal groups of actors in the AI industry: the state, the academic community, and business. The document was developed under the National Strategy for the Development of Artificial Intelligence for the period until 2030 approved by Executive Order No. 490 of the President of the Russian Federation "On Developing Artificial Intelligence in the Russian Federation" dated October 10, 2019.[56] The key objectives of the Code are to build trust in AI technologies and ensure a human-centric approach.

The code is universal and does not touch upon the military use of AI technologies. The document has two sections. The first expounds the principles and rules that reflect the human-oriented approach, including respect for human autonomy and free will, non-discrimination, knowledge of and compliance with Russian legislation, and accounting for potential risks within AI systems' life cycle. This section also dwells on the risk-oriented approach, the responsible attitude of actors (responsibility ultimately and fully rests with the human being), non-maleficence, the possibility of identifying AI while interacting with human beings, and the presumption that human beings can stop such interactions at will. Information security and data processing quality are treated separately. As regards support for the industry, technological development, strengthening competences and promoting cooperation between developers, supporting infrastructure development, expanding data accessibility and improving the quality of data labelling, and providing financial support at all levels are all prioritized over competition.

The second section focuses on the Code's legal foundations, which are based on the

[51] https://d-russia.ru/kitaj-vpervye-opublikoval-nastavlenie-po-jetike-ispolzovanija-ii-sistem.html.

[52] https://www.verdict.co.uk/china-ai-rulebook/.

53 https://www.globalgovernmentforum.com/china-unveils-ai-ethics-code/.

[54] https://freedomhouse.org/report/freedom-net/2021/global-drive-control-big-tech.

[55] Ethical Code for Artificial Intelligence. https://a-ai.ru/code-of-ethics/.

[56] Executive Order No. 490 of the President of the Russian Federation "On Developing Artificial Intelligence in the Russian Federation" dated October 10, 2019. http://www.kremlin.ru/acts/bank/44731.

Constitution of the Russian Federation and strategic documents that determine the industry's medium-term development and clearly outline the circle of actors within AI system's life cycle. Individual provisions stipulate the voluntary mechanism and procedure of acceding to the Code and introduce the mechanism of an ethics ombudsman. It is important that there is an option of developing guidelines and methodologies on applying the Code, and that there is a compendium of practices for improving the quality of interactions within the industry and maintaining optimal development that benefits all actors.

The document is expected to create a solid foundation for the long-term development of soft regulations along with documents adopted on matters of strategic development.

It is important to keep in mind that, at this stage, codification has great potential for building a ramified hierarchy of codes in all economic sectors where AI technologies are used. Moreover, some sectors may introduce further segmentation to handle problems in narrower areas. Currently, Russia has the potential to promote this approach both domestically and internationally. The most likely option is working in the EAEU, where its Digital Agenda until 2025 is already in effect,[57] and codification initiatives have a chance at becoming a good practical step in stimulating the development of the AI industry in adjacent states.

[57] Decision No. 12 of the Supreme Eurasian Economic Council "On the Principal Areas for Implementing the Eurasian Economic Union's Digital Agenda until 2025" dated October 11, 2017. November 16, 2017. https://www.garant.ru/products/ipo/prime/doc/71708158/.

Track Three: Shaping the Stance of Business on the Role of Ethics in Artificial Intelligence

In this section, we attempt to briefly outline the key aspects in the stances of private business to the development of AI-based software. This section presents an overview of AI development strategies in some companies and the current and, crucially, publicly available examples and capabilities of private businesses in AI technologies used to solve problems faced by states.

The recent phenomenon of the "metaverse" is worth mentioning here, as it will indisputably be developed and augmented in the near future, thus changing our reality and public approaches.

PRIVATE COMPANIES THROUGHOUT THE WORLD AND THEIR STRATEGIES FOR DEVELOPING AI: A WINDOW OF OPPORTUNITY FOR SETTING ETHIC FRAMEWORK IN MANAGEMENT

To open this section, we would like to briefly mention an interesting dialogue that took place between Hilary Mason, founder and CEO of Fast Forward Labs, and Jake Porway, founder and executive director of DataKind, on business support and the future of "responsible" AI. The dialogue is published on the website of the Rockefeller Foundation.[58]

It is clear from the very start of the discussion that the parties do not have a single concept of "responsible" AI. Mason stresses the need to think at the outset about the potential impact of AI on people, while technology itself, in her opinion, "has to be owned by the product leaders, the business strategists and the people making business-model decisions as much as it is owned by the technologists doing the technical work." Porway, on the other hand, when speaking about responsible AI, says that "the responsibility for any technology comes down to who has oversight of a system and who says yes or no. It depends on who can say this goes forward or not. It's funny that we automate these processes and tasks and let AIs do their thing."

They believe that private companies are only beginning to talk about introducing previously developed codes of ethics into the operations of their technological (engineering) departments, and this is a positive development. It should be noted, however, that the world has no single concept of "responsible AI," and if there is no clear understanding, it is impossible to either create a methodology, or evaluate the outcome.

Mason says that today, "unfortunately, running a rigorous experiment to determine whether your AI interviewer is better or worse than a human is really hard because a huge set of complex economic and demographic factors make it hard to assess such AI systems."

Currently, the company is working on formulating metrics (project economy) that touch upon "optimizing for profits" and "how these things correlate with numbers of sales" that would be understandable to businesses. The CIO remarks that philosophical aspects inherent in ethical principles are not easily quantified.

"Success in the real world is hard to quantify because the code is too complex and because we all have independent sets of values for how

[58] https://www.rockefellerfoundation.org/blog/taking-care-of-business-the-private-sectors-lens-on-responsible-ai/.

we think the world should be. But AI systems work only when they have very specific objective functions. The greatest trick AI will pull off will not be taking over humanity. Often, we're not explicit enough about what success looks like in society." The issue of assessing it with the use of a mathematical apparatus currently remains a big question.

It is no secret that now many companies are already investing in developing AI technologies. Researchers from *MIT Sloan Management Review,* BCG Gamma, and BCG Henderson Institute polled over 2500 business leaders in 2019 and drew an interesting conclusion.[59]

They identify a group of companies that do not merely invest in AI, but create added value for the company.

It is interesting that merely implementing AI as a technology or a set of technologies is not enough for a company to obtain maximum effect. The company only derives profit and value when the strategy of using AI-based technologies is integrated into its corporate strategy.

It is important that the very process of integrating AI into corporate strategy guarantees that AI projects slated for implementation will attract the attention of every company employee. This fact has the crucial potential effect. As the authors of the study note, "With many possible AI applications across the enterprise, AI-specific strategies that aren't aligned to the overall business strategy inevitably lead to scattered, ineffectual efforts." "Linking AI with business strategy helps companies zero in on initiatives that bring or facilitate the most important outcomes. That makes for a savvier, much more effective allocation of AI talent and resources." An important sequence of actions that is common for all leaders is answering questions such as "What are our business objectives – and how can AI help us meet them?" In answering such questions,

Companies that derive value from AI are more likely to integrate their AI strategy with their overall corporate strategy.

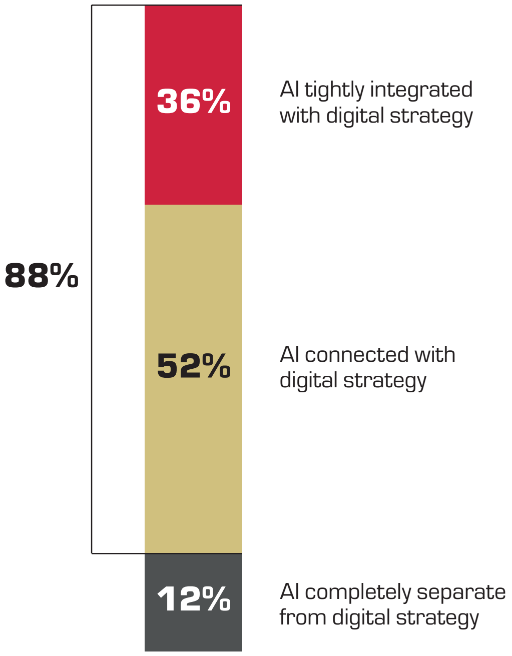

Figure 2: Different types of companies and AI value for them depending on integrating AI technologies into their corporate development strategies [60]

Organizations that have seen revenue impact from AI see the potential for more.

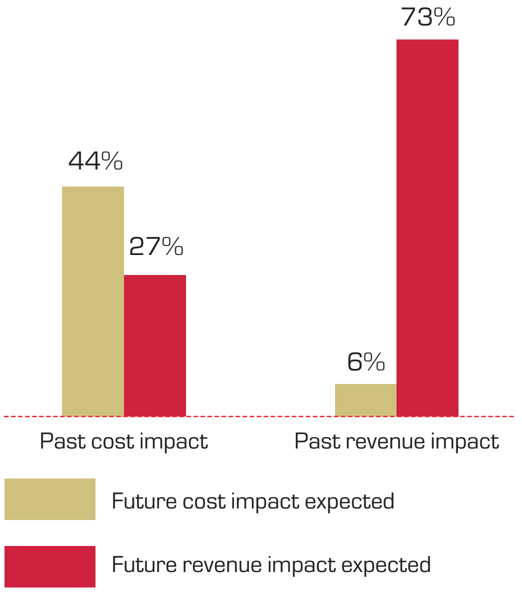

Figure 3: Organizations that have seen the value of introducing AI are more optimistic about the future value of such investment[61]

[59] https://web-assets.bcg.com/img-src/Final-Final-Report-Winning-With-AI-R_tcm9-231660.pdf.
[60] https://web-assets.bcg.com/img-src/Final-Final-Report-Winning-With-AI-R_tcm9-231660.pdf.
[61] https://web-assets.bcg.com/img-src/Final-Final-Report-Winning-With-AI-R_tcm9-231660.pdf.

companies develop logical actions in terms of automating current processes and cost-cutting, and also have the chance to project the further impact that artificial intelligence will have on their business.

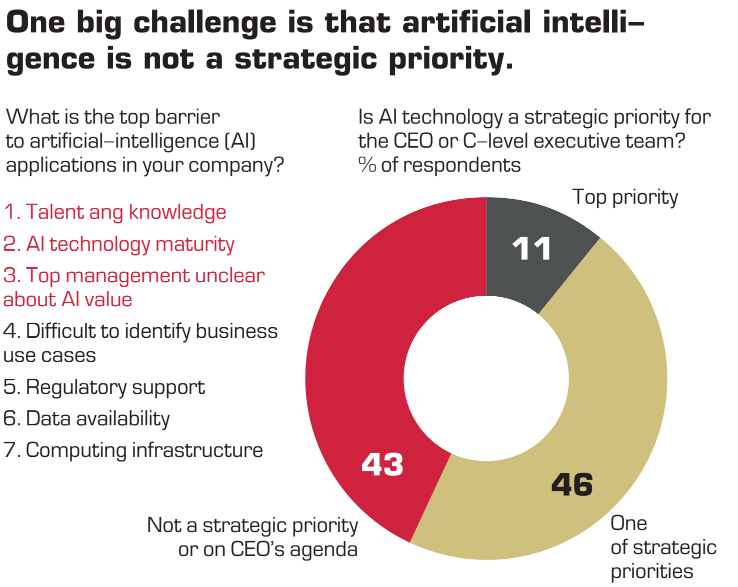

Figure 4: McKinsey & Company: one big challenge is that artificial intelligence is not a strategic priority [62]

The study examines the case of Deutsche Bank in great detail. For instance, for one of its credit products in Germany, the company used AI to make real-time decisions on approving loans, that is, by the time a client has finished filling out an online application, the software already knows whether or not the loan will be approved. This approach guaranteed German citizens that a person's credit history would not be ruined if their credit application was denied. The study's authors note that "for that specific product, loan issuance shot up 10- to 15-fold in eight months after the AI-powered service was launched." For Deutsche Bank, therefore, real profit is the opportunity to get in touch with clients who would not even have attempted to apply for a loan through a more traditional process.

In its market analysis,[63] the management consulting firm **McKinsey & Company** claims that many businesses simply do not understand how to integrate AI technologies into their corporate strategies.

"Given the buzz and confusion that surrounds AI in general, business leaders need to determine what the technology can and cannot do for their company, and build an AI strategy based on those findings."

The time-tested algorithm of action would be to take the following consecutive steps (the first two are geared towards the external environment, while the latter two are geared towards the company's internal environment):

1) identifying potential applications
2) playing out scenarios of AI-generated industry disruptions
3) defining a strategic stance and selecting underlying AI initiatives,
4) making the AI transformation happen.

METAVERSES AND ARTIFICIAL INTELLIGENCE: PROSPECTS OF AN ETHICAL CONSTRUCTION

The word "metaverse," a portmanteau of meta- (meaning "to transcend") and "universe," describes a hypothetical synthetic environment connected to the physical world. The word "metaverse" was coined by Neal Stephenson in his play *Snow Crash* (1992).[64]

Of special interest here is an article written by a team of researchers from Chinese universities published in October 2021 that attempts to shed light on all the structural characteristics of the "metaverse" concept.[65]

We find the article's attempt to present a general picture of the functioning of various

[62] https://www.mckinsey.com/industries/automotive-and-assembly/our-insights/how-advanced-industrial-companies-should-approach-artificial-intelligence-strategy.

[63] https://www.mckinsey.com/industries/automotive-and-assembly/our-insights/how-advanced-industrial-companies-should-approach-artificial-intelligence-strategy.

[64] Judy Joshua. Information Bodies: Computational Anxiety in Neal Stephenson's Snow Crash. Interdisciplinary Literary Studies, 19(1):17–47, 2017. Publisher: Penn State University Press.

[65] https://www.researchgate.net/publication/355172308_All_One_Needs_to_Know_about_Metaverse_A_Complete_Survey_on_Technological_Singularity_Virtual_Ecosystem_and_Research_Agenda

levels of the metaverse particularly important for the purposes of our overview.

The virtual world is certainly based on "hardware," the IT-infrastructure that allows the information and communications devices to run together smoothly and efficiently, and offers users access to the "metaverse."

For the metaverse infrastructure to work, eight pillars are required:

1) Network (the Internet, for instance)
2) Edge/cloud
3) Artificial intelligence
4) Computer vision
5) Blockchain
6) Robotics (IoT)
7) User interactivity
8) Extended reality

The infrastructure serves as the foundation for the metaverse ecosystem and its six pillars:

1) Avatar
2) Content creation
3) Virtual economy
4) Social acceptability
5) Security and privacy
6) Trust & Accountability.

Considering the above-listed metaverse components through the lens of business development, virtually all companies have indisputably completed the first stage of creating the basis, the IT infrastructure. It is perhaps in the context of actively developing the metaverse concept in the near future that they should adjust their digital transformation strategies somewhat, building in capabilities for greater integration with such technologies as AI, augmented and extended reality, and payment and finance instruments.

Special mention should be made of the fact that the Russian Federation is developing all the metaverse components listed above to a greater or lesser extent, and that they require a more detailed study in order to formulate recommendations for intensifying this development.

The question of the ethical development of metaverses is growing in importance in terms of developing a new level of digitization in society and economy in general.

Back in 2020, almost a year before the transnational corporation Facebook announced its Metaverse, *Forbes* published an article on the ethical aspects of the development of metaverses and privacy and confidentiality in the new digital reality.[66]

An important point was made by Kavya Pearlman, who is quoted in the article as saying, "In a new world where we extend reality and defy reality, a world where data fuels the progress we make in the metaverse, we have to hold big tech accountable for transparency and ethical use of data being collected." The article also mentions the XR Safety Initiative, i.e. principles developed jointly by tech companies and society in order to ensure trust between people and new technologies and create the possibility of building safe immersive digital ecosystems.

Speaking at the online event GamesBeat Summit: Into the Metaverse, Richard Bartle, one of the leading academics on video games and a senior lecturer and honorary professor of computer game design at the University of Essex in the United Kingdom, described his vision of the metaverse ethics.[67] The metaverse, in his opinion, should become a "collective scheme for allowing multiple 3D environments to interoperate and communicate with each other in much the same way as the internet does but it's 3D. It has reality aspects to it and virtual reality aspects to it." Special attention should be paid to curtailing toxicity or "bad" behaviour that is not accepted by society, the kind of behaviour where players harass or bully someone. Bartle insists that metaverse creators must make it safe for everyone since in the future virtually everyone will be in the metaverse. At the same time, Bartle notes that ethical norms within the metaverse, both for conduct and for its use, will change over time, and today, we can only guess at what norms will be acceptable in the future.

[66] https://www.forbes.com/sites/cathyhackl/2020/08/02/now-is-the-time-to-talk-about-ethics--privacy-in-the-metaverse/?sh=4414af-caae6c.

[67] https://venturebeat.com/2021/01/28/the-ethics-of-the-metaverse/.

CONCLUSION

AI ethics is becoming one of the most important multidisciplinary fields today. All the main groups of actors have put forward their perspectives on the problem, and their emphases depend on their accumulated experience and understanding of level of risk. Today, we are seeing a period of qualitative transition from formulating and understanding the principles of applying AI to the first practical steps in ethics, which means that the use of soft law in the industry is indeed being expanded.

This paper has considered the emerging approaches to the discussion of ethics in the leading academic schools, as well as at the national level in the context of soft law as part of state strategies, starting with codes of ethics and concepts proposed by businesses, including transnational companies. At the international level, 2021 was the year that UNESCO adopted the Recommendation on the Ethical Aspects of AI, the first framework document that addresses the long-term ethical agenda for the development of the AI industry in the long term. The medium- and long-term prospects of the industry's development clearly entail transitioning from hype and market overheating to assessing the effectiveness of local experience in AI application and understanding the ethical risks involved in scaling these practices.

Clearly, ethical issues become increasingly relevant as the metaverse concept is being popularized and disseminated. Currently, this concept is commercial and is essentially a long-term development doctrine for companies that promoting the concept. Evidently, a radically new approach to developing the AI industry may emerge in the medium term if the need for synergy and balance, including in ethics, between multilevel AI systems appears. Data transfer within metaverses is becoming more complex, since metaverses are by nature cross-border phenomena, and the understanding of data ethics and AI ethics in different ethnic groups is becoming a key factor.

Russia adopted its national ethics code in 2021. The code is a consensus document developed jointly by all the main actors involved in the AI system's life cycle. This provides us with an opportunity to further promote the codification initiative at the sectoral level.

REFERENCES

1. 1384th Meeting, 23 September 2020, 10.1 Ad Hoc Committee on Artificial Intelligence (CAHAI) [E-resource]. Available at: URL https://search.coe.int/cm/Pages/result_details.aspx?Oblectid=09000016809fa65b

2. A Human-Centric Artificial Intelligence, The Pontifical Academy for Life, [E-resource]. Available at: URL https://www.romecall.org/

3. A Practical Framework to Develop AI Responsibly. The Responsible Machine Learning Principles [E-resource]. Available at: URL https://ethical.institute/principles.html

4. A Principled AI Discussion in Asilomar, The FLI Team [E-resource]. Available at: URL https://futureoflife.org/2017/01/17/principled-ai-discussion-asilomar/

5. AI Ethics Guidelines: European and Global Perspectives, Ad Hoc Committee on Artificial Intelligence (CAHAI) [E-resource]. Available at: URL https://rm.coe.int/cahai-2020-07-fin-en-report-ienca-vayena/16809eccac

6. AI Ethics Lab [E-resource]. Available at: URL https://aiethicslab.com/about/

7. AI Ethics Principles & Guidelines, Smart Dubai [E-resource]. Available at: URL https://www.digitaldubai.ae/docs/default-source/ai-principles-resources/ai-ethics.pdf?Sfvrsn=d4184f8d_6

8. AI Policy – China. Future of Life Institute. 2018. [E-resource]. Available at: URL https://futureoflife.org/ai-policy-china/

9. AI Policy – China. Future of Life Institute. 2018. [E-resource]. Available at: URL https://futureoflife.org/ai-policy-china/

10. Now Is The Time To Talk About Ethics And Privacy In The Metaverse 2020, Forbes [E-resource]. https://www.forbes.com/ sites/cathyhackl/2020/08/02/now-is-the-time-to-talk-about-ethics--privacy-in-the-metaverse/?sh=4414afcaae6c

11. AI4People's Ethical Framework for a Good AI Society: Opportunities, Risks, Principles, and Recommendations [E-resource]. Available at: URL https://www.eismd.eu/wp-content/uploads/2019/02/Ethical-Framework-for-a-Good-AI-Society.pdf

12. The Ethics of the Metaverse [E-resource]. https://venturebeat.com/2021/01/28/the-ethics-of-the-metaverse/

13. ALL One Needs to Know about Metaverse: A Complete Survey on Technological Singularity, Virtual Ecosystem, and Research Agenda. lik-Hang Lee1, Tristan Braud, Pengyuan Zhou, Lin Wang1, Dianlei Xu6, Zijun Lin, Abhishek Kumar,Carlos Bermejo, and Pan Hui,Fellow, IEEE [E-resource] Available at: URL https://www.researchgate.net/publication/355172308_All_One_Needs_to_Know_about_Metaverse_A_Complete_Survey_on_Technological_Singularity_Virtual_Ecosystem_and_Research_Agenda

14. Analyzing Artificial Intelligence Plans in 34 Countries. Samar Fatima, Kevin C. Desouza, Gregory S. Dawson, and James S. Denford [E-resource]. Available at: URL https://www.brookings.edu/blog/techtank/2021/05/13/analyzing-artificial-intelligence-plans-in-34-countries/

15. Artificial Intelligence: Commission Takes Forward its Work on Ethics Guidelines [E-resource]. Available at: URL https://ec.europa.eu/commission/presscorner/detail/en/IP_19_1893

16. Asia's AI Agenda: The Ethics of AI. MIT Technology Review Insights, 2019. [E-resource]. Available at: URL https://www.technologyreview.com/2019/07/11/134229/asias-ai-agenda-the-ethics-of-ai/

17. China Unveils AI Ethics Code, Global Government Forum [E-resource]. Available at: URL https://www.globalgovernmentforum.com/china-unveils-ai-ethics-code/

18. China's New AI Rulebook: Humans Must Remain in Control [E-resource]. Available at: URL https://www.verdict.co.uk/china-ai-rulebook/

19. Convention for the Protection of Individuals with Regard to Automatic Processing of Personal Data [E-resource]. Available at: URL https://rm.coe.int/1680078b37

20. Current and Future Challenges of Coordinated Policies on AI Regulation: International Conference [E-resource]. Available at: URL https://www.coe.int/en/web/artificial-intelligence/-/-current-and-future-challenges-of-coordinated-policies-on-ai-regulation-international-conference

21. Ethics and Governance of Artificial Intelligence for Health. WHO Guidance [E-resource]. Available at: URL https://www.who.int/publications/i/item/9789240029200

22. Feasibility Study, Ad Hoc Committee On Artificial Intelligence (CAHAI) [E-resource]. Available at: URL https://rm.coe.int/cahai-202канада0-23-final-eng-feasibility-study-/1680a0c6da

23. Government AI Readiness Index 2020, Oxford Insights [E-resource]. Available at: URL https://static1.squarespace.com/static/58b2e92c1e5b6c828058484e/t/5f7747f29ca3c20ecb598f7c/1601653137399/AI+Readiness+Report.pdf

24. Gutierrez, Carlos Ignacio and Gutierrez, Carlos Ignacio and Marchant, Gary E., A Global Perspective of Soft Law Programs for the Governance of Artificial Intelligence (May 27, 2021). [E-resource]. Available at: URL https://ssrn.com/abstract=3855171 or http://dx.doi.org/10.2139/ssrn.3855171

25. How Advanced Industrial Companies Should Approach Artificial-Intelligence Strategy [E-resource]. Available at: URL https://www.mckinsey.com/industries/automotive-and-assembly/our-insights/how-advanced-industrial-companies-should-approach-artificial-intelligence-strategy

26. Jobin, A., Ienca, M. & Vayena, E. The Global Landscape of AI Ethics Guidelines. Nat Mach Intell 1, 389–399 (2019) [E-resource]. Available at: URL https://doi.org/10.1038/s42256-019-0088-2

27. Judy Joshua. Information Bodies: Computational Anxiety in Neal Stephenson's Snow Crash. Interdisciplinary Literary Studies, 19(1):17–47, 2017. Publisher: Penn State University Press

28. Leslie, D. (2019). Understanding Artificial Intelligence Ethics and Safety: A Guide for the Responsible Design and Implementation of AI Systems in the Public Sector. The Alan Turing Institute. [E-resource]. Available at: URL https://doi.org/10.5281/zenodo.3240529

29. National AI Strategy [E-resource]. Available at: URL https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/1020402/National_AI_Strategy_-_PDF_version.pdf

30. Newman, M. H. A. (1955). "Alan Mathison Turing. 1912–1954." Biographical Memoirs of Fellows of the Royal Society. 1: 253–263. Doi:10.1098/rsbm.1955.0019

31. OECD AI Policy Observatory [E-resource]. Available at: URL https://oecd.ai/en/

32. OECD Digital Economy Papers [E-resource]. Available at: URL https://www.oecd-ilibrary.org/science-and-technology/oecd-digital-economy-papers_20716826

33. Protocol Amending the Convention for the Protection of Individuals with Regard to Automatic Processing of Personal Data (CETS No. 223) [E-resource]. Available at: URL https://www.coe.int/en/web/conventions/full-list?Module=treaty-detail&treatynum=223

34. Rebecca Arcesati. Lofty Principles, Conflicting Incentives: AI Ethics Governance in China. Mercator Institute for China Studies. 2021. [E-resource]. Available at: URL https://merics.org/en/report/lofty-principles-conflicting-incentives-ai-ethics-and-governance-china

35. Statement on Algorithmic Transparency and Accountability, Association for Computing Machinery US Public Policy Council (USACM) [E-resource]. Available at: URL https://www.acm.org/binaries/content/assets/public-policy/2017_usacm_statement_algorithms.pdf

36. Supporting Policy with Scientific Evidence [E-resource]. Available at: URL https://knowledge4policy.ec.europa.eu/ai-watch_en

37. Taking Care of Business: The Private Sector's Lens on Responsible AI, The Rockefeller Foundation. [E-resource]. Available at: URL https://www.rockefellerfoundation.org/blog/taking-care-of-business-the-private-sectors-lens-on-responsible-ai/

38. The Beijing Artificial Intelligence Principles. Wired – The Latest in Technology, Science, Culture and Business. 2019. [E-resource]. Available at: URL https://www.wired.com/beyond-the-beyond/2019/06/beijing-artificial-intelligence-principles/

39. The Global Drive to Control Big Tech, Freedom house [E-resource]. Available at: URL https://freedomhouse.org/report/freedom-net/2021/global-drive-control-big-tech

40. The National AI Initiative and Connection Point to Ongoing Activities to Advance U.S. Leadership in AI [E-resource]. Available at: URL https://www.ai.gov

41. The Recommendation on Artificial Intelligence (AI), OECD, [E-resource]. Available at: URL https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0449

42. The Recommendation on the Ethics of Artificial Intelligence [E-resource]. Available at: URL https://unesdoc.unesco.org/ark:/48223/pf0000379920.page=14

43. Whittlestone, J. Nyrup, R. Alexandrova, A. Dihal, K. Cave, S. (2019). Ethical and Societal Implications of Algorithms, Data, and Artificial Intelligence: A Roadmap for Research. London: Nuffield Foundation.

44. Winning With AI findings from the 2019 Artificial Intelligence Global Executive Study and Research Project, BCG [E-resource]. Available at: URL https://web-assets.bcg.com/img-src/Final-Final-Report-Winning-With-AI-R_tcm9-231660.pdf

45. Younas, Ammar, Beijing Consensus of Artificial Intelligence for Children: An Effort to Prevent Juvenile Delinquency (September 19, 2020). [E-resource]. Available at: URL https://ssrn.com/abstract=3695631 или http://dx.doi.org/10.2139/ssrn.3695631

46. Bioethics and Biotechnologies: Limits for Improving the Human Being. A Festschrift for Pavel Tishchenko on his 70th Anniversary. E. G. Grebenshchikova, B. G. Yudin, eds. (in Russian). Moscow: Moscow Humanities University Press, 2017. 240 p.

47. WHO Issues First Global Report on Artificial Intelligence (AI) in Health and Six Guiding Principles for Its Design and Use [E-resource]. Available at: URL https://www.who.int/news/item/28-06-2021-who-issues-first-global-report-on-ai-in-health-and-six-guiding-principles-for-its-design-and-use

48. Digital Economy Report 2021, UNCTAD [E-resource]. Available at: URL https://unctad.org/system/files/official-document/der2021_en.pdf

49. China Published its First Instructions on the Ethics of Using AI Systems (in Russian) [E-resource]. Available at: URL https://d-russia.ru/kitaj-vpervye-opublikoval-nastavlenie-po-jetike-ispolzovanija-ii-sistem.html

50. Ethical Code for Artificial Intelligence [E-resource]. Available at: URL https://a-ai.ru/code-of-ethics/

51. Asilomar AI Principles [E-resource]. Available at: URL https://futureoflife.org/2017/08/11/ai-principles/

52. Decision of the Supreme Eurasian Economic Council of October 11, 2017 No 12 "On the Principal Areas for Implementing the Eurasian Economic Union's Digital Agenda until 2025." November 16, 2017. [E-resource]. Available at: URL https://www.garant.ru/products/ipo/prime/doc/71708158/

53. Executive Order No. 490 of the President of the Russian Federation "On Developing Artificial Intelligence in the Russian Federation" dated October 10, 2019 (in Russian). [E-resource]. Available at: URL http://www.kremlin.ru/acts/bank/44731

54. AI Ethics: Another Step Closer to the Adoption of UNESCO's Recommendation [E-resource]. Available at: URL https://en.unesco.org/news/ai-ethics-another-step-closer-adoption-unescos-recommendation-0

www.mgimo.ru

приоритет2030^
лидерами становятся

MGIMO
UNIVERSITY