



Anna Abramova
Anastasia Ryzhkova
Iuliia Tserekh

GLOBAL AI ETHICS ASSESSMENT KEY SUBINDEXES: THINK TANKS

AI FOR DEVELOPMENT research paper collection

приоритет2030⁺
Лидерами становятся

MOSCOW STATE INSTITUTE OF INTERNATIONAL
RELATIONS
(UNIVERSITY) OF THE MINISTRY OF FOREIGN AFFAIRS OF RUSSIA

MGIMO Centre For AI

Anna Abramova, Anastasia Ryzhkova, Iulia Tserekh

**“AI ETHICS ASSESSMENT AT NATIONAL AND INTERNATIONAL
LEVELS. THE APPROACH TO INDEX FRAMEWORK AND
METHODOLOGY”**

RESEARCH CENTRES AND THINK TANK

Research paper

Moscow, 2022

ISBN 978-5-6047689-6-9

Authors:

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

Anastasia Ryzhkova, PhD, researcher MGIMO Centre for AI

Iulia Tserekh, junior researcher MGIMO Centre for AI

Abstract

Artificial intelligence (AI) ethics become one of the essential elements of soft law in regulating national and international market. December 2021 UNESCO adopted the Recommendation on the ethics of artificial intelligence that provides the approaches for international soft regulation putting ethics in the heart. The Global AI Ethics index framework could be the basis for ethical impact assessment in alliance with the Recommendation and OECD AI Principles and the framework for AI classification. Methodology places the human in the center and includes all the key stakeholders along the AI system lifecycle. The data for the index could be taken from the existing databases of UNESCO, OECD, UNCTAD, WBG. But specificity of the topic enlarges on clarification and elaboration of the data that could be added to national statistics tables at macrolevel and also from the surveys, that cover the microlevel.

“AI ethics assessment at national and international levels. The approach to index framework and methodology” is the first publication in AI FOR DEVELOPMENT research paper collection

JEL F01, F20, F42, F53, F55, F60, F63, F68

Key words: artificial intelligence, ethics, index, soft law

© 2022 MGIMO. All rights reserved. Short sections of text, not to exceed two paragraphs, may be quoted without explicit permission provided that full credit, including © notice, is given to the source.

Cover photo: canva.com

Moscow, 2022

Contents

List of abbreviations	5
Introduction.....	6
Research centers landscape: the growing diversity	7
Methodology.....	8
Conclusion	16
References.....	17

List of abbreviations

AI	Artificial intelligence
ICT	Information-communication technology
IP	Intellectual property
NSF	National Science Foundation
OECD	Organisation for Economic Co-operation and Development
UNESCO	United Nations Educational, Scientific and Cultural Organization
R&D	Research and development
WEF	World Economic Forum

Introduction

This working paper is the next step in the ongoing research of MGIMO Centre for AI on elaboration of Global AI ethics index. In the separate papers all the key subindexes are discussed in details including the key indicators and challenging issues.

Global AI ethics index is a complex approach involving all the key actors along AI life cycle – state, business, civil society, research centres/ think tanks. Moreover, we propose the estimation through three subindexes of the areas that contribute AI sustainable development at all the stages and influence all the groups of actors – AI literacy, R&D investments and ICT infrastructure development.

The General Framework for the Global AI ethics index published in February 2022 (Abramova, Ryzhkova and Tserekh , 2022) covers all the main elements of the index presenting the groups of key indicators. This research paper is focused on detailed coverage of the subindex Research centers/think tanks contribution to the AI ethics perception, developments, highlights the challenging issues for all the types of research organization.

The structure of the research is follows – the first section is introduction, the second is devoted to the coverage of the key trends and issues with regard to AI development for research centers, the third one is dedicated to detailed methodology of Global AI ethics subindex focused research centers possible contributions with regard to AI ethics and the final one is dedicated to the discussion covering the most challenging issues in practical implementation of the subindex within different groups of research centers.

The authors are grateful to the leaders and coordinators of the National Priority 2030 project for making it possible to conduct the research.

Research centers landscape: the growing diversity

Research centers are one of the main pillars in AI development. The growing investments in AI push the demand for technological advances, understanding the technology's potential for improvements in efficiency, discussions on multidisciplinary issues including AI ethics. The growing research interest to AI ethics was registered through the growing number of publications sharply increased after 2015, in 2019 their number reached 70 per year¹.

Research centers could contribute all mentioned above depending on the aim and scope of their functioning.

The growing landscape of AI research activities could be classified relying on the following key characteristics – focus on AI, public or private, with regard to sectoral distribution, localization, involvement in different collaborations.

One of the first indicators to mention could be called “AI focus”. The large number of experts started their research with the centers with broad scope of activities from economy developments to digital technology. Later on, in these type of centers AI focused laboratories or initiatives with focus on AI developments emerged. One of the vivid examples is Brookings institute, nonprofit public policy organization conducting research on development at national and international levels. One of the latest initiatives from Brookings is establishing The Forum for Cooperation on Artificial Intelligence, having ethics as one of the discussion issues (Kerry C., MELTZER J., RENDA A., 2022).

The last several years were marked with new wave of growing public investments in AI in the leading countries. Public long-term support to fundamental AI research during the priviest decades provided the ground for further short-term investments from private sector (OECD 2021). During the period 2001-2019 US public funding saw 17 times increase.

¹NSF partnerships expand National AI Research Institutes to 40 states. July 29, 2021.
<https://hai.stanford.edu/news/state-ai-10-charts>

In 2020 National Science Foundation (NSF) established seven National AI Research Institutes. In 2021 the list was enlarged with new eleven centers focused on fostering AI development in collaboration with the leading private companies and federal agencies. According to NSF combined investments reach 220 \$ mln²

Besides, private business establish AI research units focused mainly on technology advances, less on AI ethics research. According to Salesforce AI ethics was one of the main concerns for consumers in 2018³. AI leaders from business responded with adoption of AI principles and introduction of administrative units or AI ethics teams⁴. In this regard subindex Business better reflects contribution from the private sector in AI ethics developments.

With regard to sectoral distribution OECD approach for analysis of AI-related R&D funding could be applied (OECD 2021). The key sectors could be “general AI techniques, AI prerequisites and impact (such as education and training and social impact), AI fields (such as computer vision and natural language processing), medical AI applications, and non-medical AI application areas (such as business and the social sciences). For research centers from each of the mentioned above sectors AI ethics metrics could be applied.

The localization metrics could cover national, regional or international levels. In this regard WIPO methodology in IP data bases could be applied.

Methodology

Review of possible scenario

The sub-index for assessing the ethical aspects of AI at the level of think tanks requires the involvement of a wide range of participants (universities, schools, scientists) or the creation of special statistics at the state level.

² <https://beta.nsf.gov/news/nsf-partnerships-expand-national-ai-research>

³ Ethical Leadership and Business. Salesforce 2018.).

https://www.salesforce.com/content/dam/web/en_us/www/documents/research/salesforce-research-ethical-leadership-and-business.pdf

⁴ <https://www.wired.com/story/tech-firms-move-to-put-ethical-guard-rails-around-ai/>

The authors got the expert approach as the basis of the assessment methodology. This approach is the most relevant for investigation issues and are on point with the current geopolitical world situation.

Firstly the authors reviewed and compared the five most common and proven methods for evaluating complex and dynamic systems, such as:– brainstorming;

- analysis of weaknesses and strengths;
- method of charting;
- Delphi method;
- expert evaluation.

Each of the overviewed methods has its own characteristics and limitations in application.

1) Brainstorming

This method is quite good for think tank-level assessments, but is not suitable for a framework approach.

As a rule, brainstorming is carried out within the project team with the possibility of involving a third-party expert in the work. An expert may have broad, or vice versa, highly specialized knowledge, which, in the opinion of the project team leader, is important in the implementation of the project.

The algorithm of method is rather simple and contains of several steps:

1. The participants make the most detailed list of parameters, that are relevant for the project
2. The paraments with least realization probability are deleted from long-list by the majority of participants.

Advantages of the method: the speed of obtaining the result, the ease of implementation of the method.

Disadvantages of the method: the quality of the analysis directly depends on the experience and outlook of the persons participating in the brainstorming session.

The possibility of applying the method for evaluating the ethical aspects of the use of AI technologies:

- requires the experience project team involving for implementing similar products,
- a high cost
- the complexity of involving relevant professionals.

2) Analysis of weaknesses and strengths

This method is effective, but has too complicated practical implementation even when using new digital technologies. Given the wide range of actors in the topic under consideration, this method is not well suited for periodic evaluation in the framework of multivariate analysis.

The method is similar to the assumption analysis method, however, the project team compiles a list of potential parameters, identifying and subsequently analyzing their weaknesses / strengths.

Advantages: detailed consideration of the Index parameters.

Disadvantages:

- the long-time realization of the method;
- excessive detail of the method;
- the quality of the analysis directly depends on the experience and outlook of the professionals involved.

The possibility of applying the method for evaluating the ethical aspects of the use of AI technologies: the project team, with insufficient experience, may miss significant parameters and aspects.

3) Charting analysis

For a reviewed level approach, this method is too expensive and difficult to find relevant specialists.

The method is carried out within the project team with the possibility of inviting an external expert. The analysis takes place in three stages:

- drawing up cause-and-effect relationships,
- creating a flowchart of the processes being implemented,
- drawing up impact diagrams.

Advantages: qualitative consideration of potential risks of projects.

Disadvantages: the implementation of the charting method requires the skills of the project team to work with this method and significant time costs.

The possibility of applying the method for evaluating the ethical aspects of the use of AI technologies: the application of this skill requires specialized competencies and experience.

4) Delphi method

This method is effective in the case of using digital technologies, but for the generally accepted framework for assessing the ethical aspects of AI, the time range is too long. To evaluate the activities of research centers, the method is quite good for use.

The Delphi method involves conducting a large anonymous survey of external and internal experts, summarizing the collected data, issuing completed questionnaires to another expert group, followed by a face-to-face results discussion, and then re-conducting an anonymous survey with summing up the final results and compiling a list of potential risks.

Advantages: high-quality study.

Disadvantages: the method requires the long-time realization and financial resources for implementation.

The possibility of applying the method for evaluating the ethical aspects of the use of AI technologies: the method requires a lot of time and money.

5) Method of expert assessments

The method of expert assessments is similar to the Delphi method, however, it involves an open survey of experts with experience both in the field of research centers and in the field of artificial intelligence.

The method of expert assessments is similar to the Delphi method, however, involves an open survey of experts.

Advantages: a qualitative study of the identification of potential risks.

Disadvantages: it is required creation of a base of experts who are ready to participate in a large sur.

The possibility of applying the method for evaluating the ethical aspects of the use of AI technologies: the method requires a lot of time.

Calculation formula

The authors based the assessment of groups of indicators on the significance index, which is calculated by the formula:

$$r_{ij}^k = \alpha_{ij} \beta_{ij}^k, \quad (1)$$

where

r_{ij}^k - the significance of the i-indicator, assessed by the j-th respondent, in terms of the impact on the k- factor,

$i = (1 \dots N)$, where N is the number of parameters considered in the study,

$j = (1 \dots n)$, where n is the number of responses received,

$k = (1...5)$, where 1...5 are the numbers of influence groups, respectively (respectively, cost, execution time of IT project, product quality, environment, security),

α_{ij} - the weight of the significance of the indicator i, estimated by the j-th respondent,

β_{ij}^k - the value of the "effect" of the influence of the indicator on the considered stakeholder and/or the goals pursued by him.

To assess the average value of indicators, the Index of Significance of the indicator is calculated by the formula:

$$R_i^k = \frac{\sum_{j=1}^n r_{ij}^k}{n} = \frac{1}{n} \sum_{j=1}^n \alpha_{ij} \beta_{ij}^k \quad (2)$$

Proposed parameters for calculation

AI ETHICS' INTELLECTUAL PROPERTY REGISTRATION:

- number of AI ethical cases implementation in research centers and think tanks,
- dynamic of AI ethical cases implementation (current/previous)
- project scope
- project cost
- number of patents
- dynamic of new patents' quantity
- number of involved participants

- project level: local, national, transnational
- number of AI projects without AI Ethics aspects implementation created by research centers and think tanks.
- number of AI projects without AI Ethics aspects implementation involving research centers and think tanks.
- compliance with the ethical requirements of UNESCO

ETHICAL ISSUES IN R&D:

- number of AI ethical cases accidentally noticed by research centers and think tanks' staff,
- number of appeals to the competent authorities to resolve problems
- sector /-s where AI ethical case happened
- protentional negative development scenarios
- number of real cases protecting again negative consequences of ethical aspects of AI

PUBLICATIONS ON AI ETHICS:

- number of peer-reviewed scientific journals with publications on AI ethic aspects
- dynamic of quantity of peer-reviewed scientific journals with publications on AI ethic aspects
- number of publications about AI ethical aspects in peer-reviewed scientific journals
- dynamic of quantity of publications about AI ethical aspects in peer-reviewed scientific journals
- the format of the scientific journal
- Journal audience coverage
- compliance with the ethical requirements of UNESCO

EVENTS HIGHLIGHTING AI ETHICS:

- a group of indicators covering national and international conferences and forums on the ethics of AI, which can provide all stakeholders with the opportunity to exchange views on issues for further research.

- number of popular events in the field of AI ethic aspects:

- local
- national
- worldwide
- dynamic of quantity of events
 - local
 - national
 - worldwide
- number of participants
- attendance fee
- accessibility of the event to a wide range of people
- professional level of participants
- compliance with the ethical requirements of UNESCO

RESEARCHERS WORKING WITH AI ETHICAL ISSUES:

- number of researchers working with AI ethical issues
- dynamic of researchers working with AI ethical issues (current/previous)
- project scope
- project cost
- number of involved participants
- level: local, national, transnational
- project compliance with the ethical requirements of UNESCO

Conclusion

AI focused research centers are growing in number being funded from public and private resources. National AI policies and priorities place AI ethics in the heart of midterm AI sector development. Multidisciplinary research is now attracting more and more attention both from computer scientists and social ones. One of the best indicators to reflect current stage of development are publications and events. But further improvements in data collection on IP registration are needed for restructuring the discussion on AI ethics research, especially in the field of patenting and copyright protection.

Current data highlights the growing interest to AI ethics from state agencies and business. But the data on AI ethics research is still scarce and fragmented. The subindex Research centers/think tanks indicators collection in terms of international cooperation could be supported by UNESCO and WIPO with introduction of additional tables focused on AI ethics estimations.

References

Abramova, Anna and Ryzhkova, Anastasia and Tserekh, Iuliia, AI Ethics Assessment at National And International Levels. The Approach to Index Framework and Methodology (February 18, 2022). [E-resource]. Available at SSRN: <https://ssrn.com/abstract=4096669> or <http://dx.doi.org/10.2139/ssrn.4096669>

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-closeradoption-unescos-recommendation-0>

EUROPEAN COMMISSION 2021. AI watch index. [E-resource]. Available at: URL <https://op.europa.eu/en/publication-detail/-/publication/15568192-a95f-11eb-9585-01aa75ed71a1/language-en/format-PDF/source-209026200>

Kerry C., Meltzer J., Renda A. (2022). AI cooperation on the ground: AI research and development on a global scale. Report, November 4, 2022. [E-resource]. Available at: URL <https://www.brookings.edu/wp-content/uploads/2022/11/FCAI-October-2022.pdf>

Multistakeholder group discusses ten building blocks towards creating inclusive AI policies. [E-resource]. Available at: URL <https://en.unesco.org/news/multistakeholder-group-discusses-ten-building-blocks-towards-creating-inclusive-ai-policies>

Multistakeholder group discusses ten building blocks towards creating inclusive AI policies. [E-resource]. Available at: URL <https://en.unesco.org/news/multistakeholder-group-discusses-ten-building-blocks-towards-creating-inclusive-ai-policies>

OECD 2021. Izumi Yamashita, Akiyoshi Murakami, Stephanie Cairns, Fernando Galindo-Rueda. Measuring the AI content of government-funded R&D

projects: A proof of concept for the OECD Fundstat initiative. [E-resource]. Available at: URL <https://www.oecd-ilibrary.org/science-and->

technology/measuring-the-ai-content-of-government-funded-r-d-projects_7b43b038-en

OECD 2022. FRAMEWORK FOR THE CLASSIFICATION OF AI SYSTEMS [E-resource]. Available at: URL <https://www.oecd-ilibrary.org/docserver/cb6d9eca-en.pdf?expires=1646822229&id=id&accname=guest&checksum=0D6117C31817EA6B9FFDB7A65AACAFCF>

UNESCO 2021. Recommendation on the ethics of artificial intelligence. [E-resource]. Available at: URL <https://unesdoc.unesco.org/ark:/48223/pf000038>

WIPO 2019. WIPO Technology Trends 2019. Artificial Intelligence. [E-resource]. Available at: URL https://www.wipo.int/edocs/pubdocs/en/wipo_pub_1055.pdf



MGIMO Centre for AI was established to enhance international cooperation and support collaboration with all the actors of digital economy both at national and international levels. Our multidisciplinary research is focused on international cooperation agenda, national policies for AI and business opportunities. International trade and trade policy (prioritising digital trade), sustainable development, AI ethics are the key areas of our activities.

On the basis of MGIMO-University we promote an international AI expert platform with regular conferences and round tables, peer-reviewed articles and research papers. Our enlarging network of strategic partnerships makes it possible to provide AI consulting and policy solutions both for business and government agencies.

The Centre was founded in October, 2021

Our contacts



143007, Moscow Region, Odintsovo,

Novo-Sportivnaya street, 3

<https://aicentre.mgimo.ru>

E: aicentre@inno.mgimo.ru

P: +7 903 623-95-15



<https://t.me/aicentremgimo>



приоритет2030[^]
Лидерами становятся