ETHICAL ASPECTS OF THE USE OF ARTIFICIAL INTELLIGENCE IN EDUCATION: INSTITUTIONAL, PEDAGOGICAL AND STUDENT LEVELS

CZU: 174.7:[37.013:004.8]
DOI: https://doi.org/10.53486/tids2026.07

MIHAIL CROITOR
Moldova State University, Chisinau, Republic of Moldova
e-mail: mihail.croitor@usm.md
ORCID ID: 0009-0004-5046-0811

TUDOR BRAGARU
Moldova State University, Chisinau, Republic of Moldova
e-mail: tudor.bragaru@usm.md
ORCID ID: 0000-0001-6356-2906

Abstract. The article analyzes the ethical risks of using artificial intelligence (AI) in education, taking into account the current shift towards the use of generative models. The emphasis is on locally deployed large language models. This approach helps to reduce the risks of personal data leakage to external services. However, this does not eliminate the fundamental problems of fairness, comprehensibility of decisions, accountability, and the impact of AI on human subjectivity in education (the ability of a person to independently make and justify decisions).

Methodologically, the work is based on a qualitative analysis of the framework international normative documents *UNESCO* (1; 2; 3; 4), *Montreal Declaration* (5), politicians at the state level, such as *UE AI Act, USA* (6)*Insights and Recommendations* (7)and other relevant research literature on educational measurements, algorithmic justice, generative AI, and the transformation of assessment practices.

The key result is to identify three levels of ethical risks: institutional (administration), professional (teachers) and educational cognitive (students). At the institutional level, issues of accountability, bias and lack of transparency in decisions are critical. At the teacher level, the risks of automated rejection of professional judgment, loss of skills and the use of unverified model answers increase. At the student level, cognitive atrophy, unjustified trust in plausible but unreliable answers, weakening of social learning, and growing inequality of competencies become key.

The practical contribution of the article is the definition of a structured set of principles for the ethical implementation of AI in educational organizations: institutional accountability, mandatory human control in decisions with serious consequences, the development of critical AI literacy and the design of educational practices that preserve the intellectual autonomy of students.

Keywords: artificial intelligence, education, ethics of AI, generative AI, cognitive autonomy, algorithmic justice
JEL Classification: I20, O33, K23

INTRODUCTION

The integration of AI into education is developing faster than the mechanisms of its legal and ethical regulation. There are two main approaches to the use of AI in education in the world: regulatory (hard-law, e.g. UE AI Act (6), Romania (8)) and recommendatory (soft law, e.g. USA (7), Russia). Often, automated learning and assessment systems, academic risk forecasting tools, recommendation mechanisms, and generative language models are already built into educational practice to improve the efficiency of individual processes. However, all these tools simultaneously transfer the problems typical of AI systems to the educational environment: *algorithmic bias, non-*

transparency of conclusions, ambiguous responsibility for mistakes, pressure on human autonomy. (1; 9; 10)

In recent years, some organizations have moved to the local deployment of large language models. This approach is motivated by the requirements of data protection and digital sovereignty. However, reducing the risk of external data leakage is not equivalent to the ethical security of the system as a whole. The local model preserves the inherited shifts of basic learning, can produce factually incorrect content, as well as change the structure of educational activities and the distribution of responsibility between participants in the educational process. (9; 10; 11; 12)

At the level of the international framework (UNESCO, Montreal Declaration, UE AI Act, USA Insights and Recommendation, ISTE Standards for AI in Education, etc.) the principles of Fairness, Transparency, Accountability have already been enshrined, Human-in-the-loop, Human-centered AI in education and the preservation of human dignity. However, in the practical policy of educational institutions, these principles are often applied fragmentarily. Hence the research task arises: to translate general ethical and legal provisions into an operational framework that is sensitive to the roles of specific participants in the educational system. (1; 4; 5)

The purpose of the article is to systematically analyze the legal and ethical problems of using AI in education for three groups of participants (administration, teachers, students) and to formulate the principles of implementation applicable in the context of both cloud and local use of generative AI based on large language models (LLM).

1. MATERIALS AND METHODS

This study is a qualitative thematic synthesis of XX sources, grouped into categories:

1. Normative and framework documents of the international level: recommendations and guidelines of UNESCO, OECD, WEF, on AI in education and ethics of AI, as well as the Montreal Declaration. (13; 5; 14; 2; 3; 4; 1)
2. Legislative and policy documents that define the principles of using AI in digital education. (8; 6; 7; 15)
3. University Policies, Practical and Methodological Guides. (16; 17; 18; 19)
4. Academic publications and research papers (citation of sources) on educational measurements, algorithmic fairness, explainability, the impact of generative AI on writing and critical thinking, and the effects of assessment transformation. (10; 20; 21; 22; 23; 24; 9; 25; 26; 27)

The analysis was carried out in three stages:

1. Highlighting recurring ethical problems in literature.
2. They are grouped by levels of responsibility (institutional, pedagogical, educational).
3. Comparison of the resulting groups with the principles of the international ethical framework.

Such a design makes it possible to separate the technological risks of integrating AI in education from managerial and pedagogical ones, as well as to distinguish between short-term operational effects and long-term consequences for the quality of education.

2. RESULTS AND DISCUSSION

2.1. Ethical problems of the managerial level

At the level of the educational organization, the key risk is related to the fact that it is the institution that remains the ultimate bearer of legal and moral responsibility for decisions made with the support of AI (3; 10; 9). This is especially important in scenarios with serious consequences, such as selection of applicants, transfer decisions, early warning systems for expulsion and other high-stakes decisions, such as graduation, qualification exams (28; 10; 12). For the administration of an educational organization, the following ethical problems can be distinguished:

1. Accountability and responsibility;
2. Algorithmic bias;
3. Opacity and explainability;
4. Environmental sustainability.

Accountability and responsibility. Even if a particular model is developed by an external provider, responsibility for the consequences of its application in the educational process cannot be fully delegated. Ignorance of the dataset on which the model was trained does not exempt the institution from liability. The practice of an educational organization should include clearly defined contours: who approves the model for use, who checks its quality, who considers appeals, and who makes the final decision in a conflict between the recommendation of AI and the professional judgment of a person. (3; 9; 29; 28)

Algorithmic bias. Bias do not disappear when AI is deployed locally. The on-premises infrastructure changes the trajectory of the data, but it does not invalidate the statistical and social biases inherited during the pre-training phase of the base model. For an educational organization, this means the need to regularly check for bias in groups of students and a ban on using the system in decisions with serious consequences if uneven accuracy is identified. (29; 10; 12)

Opacity and explainability. Black box decisions limit the right to a fair hearing and appeal. In the educational context, the lack of clarity has not only legal, but also pedagogical significance: if the student and the teacher do not understand the grounds for the conclusion, the decision ceases to be educationally significant. (10; 22)

Environmental sustainability. Local computational loops for large language models create additional power consumption and infrastructure overhead, which conflicts with sustainability goals and environmental, social, and governance responsibility commitments. Ethical policies for AI implementation should take into account not only learning outcomes but also the environmental cost of supporting models. (9)

Ignoring these problems can carry reputational risks for educational organizations. In summary, for the administration, the ethical problem is formulated as the need to move from the logic of "implementing a tool" to the logic of "building a system of responsible AI management".

2.2. Ethical problems for the teacher

For a teacher, AI is both a resource and a source of professional risks. Literary sources highlight the following problems:

1. Automated refusal of professional judgment;
2. Loss of skills;
3. Reliability of keeping;
4. Evaluation crisis.

Automated refusal of professional judgment. The risk of automatic abandonment of professional judgment arises when recommendations or assessments of AI are perceived as definitive. In this mode, the teacher loses the role of a responsible expert and actually confirms the decisions of the system without pedagogical interpretation. (3; 9)

Loss of skills. With the long-term delegation of routine and then complex cognitive tasks to AI, erosion of pedagogical expertise can occur: sensitivity to the quality of the student's argumentation, to the context of the group, to informal signs of academic progress decreases. (9)

Reliability of keeping. Generative AI models are able to produce a convincing but factually incorrect text. This creates an additional ethical obligation for the teacher to check, since the use of AI without verification procedures transfers the risk of error to the student and undermines trust in the educational environment. (4; 9)

Evaluation crisis. If the final assessment increasingly reflects the quality of the student's interaction with AI, rather than their own level of competency, the informative value of the assessment decreases. In the long term, this weakens the selective and motivational functions of the education system. (21; 25; 20; 30)

Thus, the main challenge for the teacher is not to abandon AI, but to maintain professional subjectivity and personal responsibility when using it.

2.3. Ethical Problems for Students

When AI is used at the learner level, the risks associated with cognitive development and the social nature of learning become central. The following risks can be distinguished:

1. Cognitive atrophy and weakening of "slow thinking";
2. Unwarranted confidence in a plausible conclusion;
3. Erosion of social learning;
4. Digital Competence Gap.

Cognitive atrophy and weakening of "slow thinking". When intellectual effort is systematically transferred to the model, the mechanisms through which critical analysis, argumentation, and the ability to solve problems are formed are weakened. This is especially important for the early stages of mastering complex disciplines. (2; 9; 11)

Unwarranted trust in a plausible conclusion. Large language models form answers based on probabilistic patterns of language, which can create the illusion of understanding and truth. For a student without developed AI literacy, the plausibility of a text is often misinterpreted as the reliability of knowledge. (2; 4)

Erosion of social learning. Hyper-personalized student-model work can reduce the intensity of interaction with the teacher and peers. As a result, the development of communication, collaborative, and socio-emotional skills that are not reducible to an individual academic result is weakened. (4; 9)

Digital Competence Gap. Even with formally equal access to AI tools, the results differ depending on the level of critical digital literacy and the quality of strategies for interacting with the model. This creates a new type of inequality: not in access to technology, but in the ability to use it without cognitive and ethical losses. Consequently, the ethically significant unit of analysis is not the fact of the student's use of AI, but the quality of the educational design in which this use occurs. (12)

Based on the risks considered at the institutional, pedagogical and student levels, taking into account the national and local context, training organizations can easily formulate a set of practical principles for the implementation of AI.

2.4. The Multidimensional Nature of Problems

According to sources, the listed problems are multidimensional in nature and are most often defined as socio-technical risks (9; 23; 27; 22). This means that the technical characteristics of AI systems (algorithms, data) are inextricably linked to specific social contexts and human values. Therefore, such problems develop from purely technical to ethical in the following cases: (9; 27; 22; 29)

1. Fundamental ethical level

Many problems are recognized as ethical, since they affect the basic philosophical categories: (9; 26; 5)

- **Deontology (rights and obligations):** Issues such as bias or digital surveillance are ethical because they violate the universal human rights to dignity, autonomy and privacy. (9; 12; 26; 5; 7)
- **Consequentialism (consequences):** The ecological footprint (energy expenditure) and "cognitive atrophy" of students are recognized as ethical problems, as they lead to long-term harm to society and future generations. (9; 10; 20)

2. Specific level of ethics of education

The studied sources draw a line between the general ethics of AI and the ethics specific to schools and universities (1; 9; 26; 10). The choice of pedagogical methods, for example, the replacement of active human mentoring with automated learning, is an ethical problem of education, as it concerns the goals of education and the very understanding of what it means to be an "educated person". (1; 2; 9; 11)

3. Intersection with other disciplines

Many of the problems mentioned above are of a "dual nature", described below:

1. **Technical and psychometric:** The issues of validity and reliability of AI assessments seem to be mathematical problems, but in education they become ethical, since any error of the algorithm directly affects the life trajectory and capabilities of the student. (23; 30; 10)

2. **Legal and regulatory:** Consent to data processing or liability for AI errors are governed by laws (such as the GDPR or the EU AI Act), but these laws are only a form of practical implementation of ethical principles. (1; 4; 9; 28; 6; 8; 7; 16; 15)

3. **Political and economic:** The dependence of educational institutions on the infrastructure of big tech companies ("digital colonialism" or "hidden commercialization") is a matter of political economy, but its ethical essence lies in the threat to education as a public good. (9; 28; 29; 14)

As a result, any technical or pedagogical failure in an AI system becomes ethical as soon as it threatens human well-being, autonomy, or fairness. (9; 27; 26; 10)

Based on the analysis, six interrelated principles for the use of AI are proposed:

1. **The principle of institutional accountability.** An educational organization should have a formalized contour of responsibility for the life cycle of an AI system: from the selection of tools to appeals against controversial decisions.
2. **The principle of mandatory human control.** In decisions with serious consequences, AI recommendations cannot be the final basis for the decision. The final check is done by a human.
3. **The principle of fairness and regular audit.** Periodic checks of the model for differentiated accuracy and discrimination effects by relevant groups are necessary.

4. **The principle of explainability sufficient for educational practice.** The system should provide explanations that allow the teacher and student to understand the grounds for the conclusion and, if necessary, challenge it.
5. **The principle of maintaining independent thinking.** Learning assignments and assessment formats are designed so that AI enhances learning, rather than replacing a key intellectual effort.
6. **The principle of critical AI literacy.** The ethical implementation of AI is impossible without systematic training of teachers and students in the limitations of models, verification practices, and the rules of academic integrity.

These principles are compatible with international regulatory frameworks, but are expressed in an operational form applicable to the policies of the educational organization. (1; 2; 3; 4; 6; 5)

2.5. Possible Strategies for Solving Ethical Problems

To solve the ethical problems that arise during the introduction of AI in education, the articles propose an integrated approach that combines **state regulation, institutional management**e, technical methods of protection and **pedagogical transformation** (1; 9; 12). At the heart of all four proposals is a humanistic approach that requires AI to remain under human control (the concept of "human-in-the-loop") and serve to augment human capabilities, not replace them. (1; 4; 3; 9)

Below are the main problem-solving strategies proposed in the literature.

1. State and regulatory regulation

Creation of a legislative framework: States are encouraged to adopt or adapt data protection laws (such as GDPR) and develop ethical frameworks for AI, translating abstract principles into legally binding norms. (1; 4; 13; 9)

Age restrictions: To protect the psyche and cognitive development of children, it is proposed to introduce age thresholds (for example, 13 years) for the independent use of generative AI. (2)

Protection of children's rights: The use of AI must comply with the UN Convention on the Rights of the Child, excluding the commercial exploitation of student data and its transformation into objects for profiling. (9)

2. Institutional Management and Audit

Institutional Policies, Guidance Materials, Guidelines, Validation and Whitelisting: Educational institutions should ensure the regulation and control of the use of AI in educational processes, conduct a preliminary ethical and pedagogical review of AI tools before their implementation, creating lists of approved and safe technologies. (4; 3; 9; 8; 15; 7)

Risk assessment and audit: It is recommended to introduce mandatory human rights impact assessment procedures and regularly audit algorithms for bias and transparency. (9; 17; 16; 6; 7; 18; 26)

Responsibility of institutions: The administration must take full legal and moral responsibility for any decisions made with the support of AI, especially in "high stakes" situations (enrollment, grading of final or certification exams). (1; 3; 9; 7; 6)

3. Technical Bias Mitigation

Sources distinguish three stages of technical solution to the problem of bias: (26; 12)

At the stage of preparation (Pre-processing): Using representative datasets, balancing samples, and adjusting weights to ensure equal representation of different groups of learners. (12; 26; 10)

In the process of learning (In-processing): Embedding ethical constraints directly into the learning algorithm so that the model minimizes discriminatory outcomes. (26; 12)

Post-processing: Adjusting AI outcome thresholds to ensure equality of opportunity between different groups of students. (26)

Pipeline of agents: In addition, a useful tool that has an impact on reducing the hallucination of large language models is the use of agent pipelines: namely, the interaction of multiple exemplars of models (agents) with well-defined roles and skills, under the control of the control agent (14).

4. Pedagogical Transformation and Protection of Subjectivity

Revision of the assessment system: To avoid "grade inflation" and plagiarism, it is proposed to move to assessing skills that AI cannot imitate (critical thinking, creativity, social context) and use multi-level assessment models such as CAIAF. A paradigm shift from evaluating a product (e.g., an essay) to critically evaluating the process and the student's skills in using AI to write, verify, and validate facts, sources, etc. (4; 25; 11; 20; 18; 14; 19; 17) (21; 20; 22; 26; 30)

Protecting the Instructor Role: Technology should complement and not replace the teacher. Teachers need to maintain autonomy in making final pedagogical decisions and act as mentors. (1; 3; 9; 10; 14; 18)

Emphasis on Social Learning: It is recommended to limit the use of AI where it deprives students of the opportunity for social interaction and collective knowledge building. (2; 9; 16; 17; 18; 19)

5. Formation of competencies (AI Literacy)

Training for all users: it is proposed to implement competency frameworks (AI CFS for students and AI CFT for teachers) not only for technical skills, but also for the development of ethical thinking. (2) (3) (1)

Critical attitude: Training should focus on demystifying AI and developing skills to verify content created by "stochastic parrots". (4; 2; 3)

6. Co-design

Co-design: it is proposed to involve teachers, students and researchers in the process of creating AI systems at the earliest stages ("ethics-by-design") so that the tools meet the real needs of the classroom. It is also important to exchange best practices between the participants in the process. (1; 27; 22; 6)

CONCLUSION

The analysis shows that the ethical problems of AI in education are multi-layered and are not reduced to the issue of data protection. Even with the local use of large language models, critical challenges remain: institutional and personal responsibility for decisions, bias and opacity of algorithms, the risks of loss of skills among teachers and students, as well as the possible weakening of independent thinking and the social component of learning among students.

The main practical conclusion is that *the sustainable adoption of AI requires not pinpoint bans or technological optimism, but a governance architecture where technological, pedagogical, and ethical decisions are aligned*.

The limitations of the work are related to the overview nature of bibliographic research and analysis and the lack of empirical measurement of effects in a particular educational organization. This opens up areas for further research:

1. Empirical assessment of the impact of AI on the quality of educational results in various disciplines;
2. Long-term studies of changes in students' cognitive strategies with the systematic use of large language models;
3. Development and validation of institutional metrics for the ethical use of AI systems in education.

REFERENCES

1. **UNESCO.** *AI and education: Guidance for policy-makers.* Paris : UNESCO, 2021. 978-92-3-100447-6.
2. —. *AI competency framework for students.* Paris : UNESCO, 2024. 978-92-3-100709-5.
3. —. *AI competency framework for teachers.* Paris : UNESCO, 2024. 978-92-3-100707-1.
4. —. *Recommendation on the Ethics of Artificial Intelligence.* Paris : UNESCO, 2021.
5. **MONTREAL DECLARATION FOR A RESPONSIBLE DEVELOPMENT OF ARTIFICIAL INTELLIGENCE.** *Montreal Declaration for a Responsible Development of Artificial Intelligence.* Montreal : Universite de Montreal, 2018.
6. **EUROPEAN PARLIAMENT and COUNCIL OF THE EUROPEAN UNION.** *Regulation (EU) 2024/1689 laying down harmonised rules on artificial intelligence (Artificial Intelligence Act).* Luxembourg : Publications Office of the European Union, 2024.
7. **U.S. DEPARTMENT OF EDUCATION, Office of Educational Technology.** *Artificial Intelligence and the Future of Teaching and Learning: Insights and Recommendations.* Washington, DC : U.S. Department of Education, 2023.
8. **GUVERNUL ROMANIEI.** *Strategia Nationala in domeniul inteligentei artificiale 2024-2027. Hotararea nr. 832/2024.* Bucuresti : Guvernul Romaniei, 2024.
9. **HOLMES, Wayne.** *Artificial Intelligence in Classrooms: Ethical Dimensions.* Brussels : European Parliament, 2026.
10. **BULUT, Okan.** *The Rise of Artificial Intelligence in Educational Measurement: Opportunities and Ethical Challenges.* s.l. : arXiv, 2024.
11. *Will AI Undermine or Support Writing and Critical Thinking?* **WILSON, Joshua.** 2, 2025, Social Innovations Journal, Vol. 30.
12. **YIN, Zhipeng.** *FairAIED: Navigating Fairness, Bias, and Ethics in Educational AI Applications.* s.l. : arXiv, 2024.
13. **EUROPEAN COMMISSION.** *Digital Education Action Plan (2021-2027).* Brussels : European Commission, 2020.
14. **OECD.** *OECD Digital Education Outlook 2026: Exploring effective uses of generative AI.* Paris : OECD Publishing, 2026.
15. **MINISTERUL DEZVOLTARII ECONOMICE SI DIGITALIZARII AL REPUBLICII MOLDOVA.** *Cartea Alba privind inteligenta artificiala si guvernanta datelor.* Chisinau : MDED, 2024.
16. **ACADEMIA DE MUZICA, TEATRU SI ARTE PLASTICE.** *Ghid privind utilizarea inteligentei artificiale.* Chisinau : AMTAP, 0.
17. **ERASMUS UNIVERSITY ROTTERDAM.** *Generative AI in Education.* Rotterdam : Erasmus University Rotterdam, 0.
18. **HARVARD UNIVERSITY INFORMATION TECHNOLOGY.** *Generative AI Guidelines.* Cambridge, MA : Harvard University, 2026.
19. **UNIVERSITATEA DE VEST DIN TIMISOARA.** *Regulament privind utilizarea inteligentei artificiale generative la UVT.* Timisoara : Universitatea de Vest din Timisoara, 2026.
20. **CHIRIKOV, Igor.** *Artificial Intelligence and Grade Inflation.* Berkeley, CA : Higher Education Working Paper Series, University of California, Berkeley, 2026.

21. *Automated Essay Scoring.* **DIKLI, Semire.** 1, 2006, Turkish Online Journal of Distance Education, Vol. 7. ISSN 1302-6488.
22. *The possibilities and limits of XAI in education: A socio-technical perspective.* **FARROW, Robert.** 2, 2023, Learning, Media and Technology, Vol. 48, pp. 266-279.
23. *Beyond Automation: A Conceptual Framework for AI in Educational Assessment.* **FARTUSNIC, Rares, ISTRATE, Olimpius and FARTUSNIC, Ciprian.** 1, 2025, Journal of Digital Pedagogy, Vol. 4, pp. 83-102. ISSN 3008-2021.
24. **FLORIDI, Luciano.** *The Ethics of Artificial Intelligence: Principles, Challenges, and Opportunities.* Oxford : Oxford University Press, 2024.
25. *Comprehensive AI assessment framework: Enhancing educational evaluation with ethical AI integration.* **KILINC, Selcuk.** 4, 2024, Journal of Educational Technology and Online Learning, Vol. 7, pp. 521-540.
26. **KIZILCEC, Rene and LEE, Hansol.** Algorithmic fairness in education. *The Ethics of Artificial Intelligence in Education.* London : Routledge, 2023, pp. 174-202.
27. **RIORDAN, Alfredo.** *Human-Centred Learning Analytics and AI in Education: A Systematic Literature Review.* s.l. : arXiv, 2024.
28. **WILLIAMSON, Ben.** *Risk, Retention, and the Algorithmic Institution*. 2024.
29. *What's the Problem with Learning Analytics?* **SELWYN, Neil.** 3, 2019, Journal of Learning Analytics, Vol. 6, pp. 11-19.
30. *A Neural Approach to Automated Essay Scoring*. **TAGHIPOUR, Kaveh and NG, Hwee Tou.** 2016. Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing. pp. 1882-1891.