Computer science

UDC 17.02:004.8

**Kamysheva Maryna**
*Senior Front End Developer*
*Company LegalMation*
*ORCID: 0009-0000-3663-5102*

# ETHICAL CHALLENGES IN THE APPLICATION OF ARTIFICIAL INTELLIGENCE IN DECISION-MAKING

***Summary.*** *Introduction. In today's rapidly evolving world of technological advancement, the use of artificial intelligence (AI) in decision-making is becoming increasingly widespread and has an impact on all areas of life, placing a special responsibility on algorithm developers and other stakeholders to ensure transparency, fairness, and accountability of the solutions adopted by algorithms. AI enables rapid analysis of large amounts of data and generation of well-founded decision-making scenarios. However, these algorithms contain hidden sources of bias, potential errors, and lack of understanding of the solutions adopted, which poses a significant threat to human rights, equality, and justice, and necessitates the development of new approaches to ensuring the ethical use of algorithms in decision-making.*

*Purpose. The purpose of the study is to determine the ethical aspects of the use of artificial intelligence in decision-making and to develop practically oriented recommendations designed to ensure transparency, controllability and a socially acceptable level of use of algorithms.*

*Materials and methods. The article uses a complex of general scientific and specialized methods of analysis and synthesis. The research materials were legislative acts, industry standards, official reports of leading research in the*

*field of artificial intelligence, as well as publications in specialized scientific journals and monographs.*

*The main methods were: analytical - to understand the state of the problem and identify the most acute ethical risks of using decision-making algorithms; synthetic - to develop practically oriented principles for ensuring transparency and controllability of algorithms; systemic - to consider algorithmic systems in the context of human rights, legislative norms and social values; and comparative - to compare the studied algorithms and develop the most acceptable decision-making models taking into account the requirements of transparency and fairness. The application of these methods allowed us to formulate conclusions and develop practically valuable recommendations for solving the identified problems.*

*Results. The scientific article identifies the main sources and manifestations of ethical problems in the application of decision-making algorithms using artificial intelligence, in particular, the lack of transparency of algorithmic models, the possibility of algorithmic bias, and the absence of control mechanisms. As a result of the study, practically oriented principles for ensuring transparency and control of algorithms have been developed, which will make it possible to improve the accuracy of decisions made and strengthen the level of trust from society. These principles are designed to ensure equal access to decision-making, minimize possible errors, and ensure a fair distribution of results among all stakeholders.*

*Discussion. In further scientific research, it is advisable to pay special attention to the development of mechanisms for ensuring the accountability of algorithms and expanding human opportunities to intervene in decision-making formed by artificial intelligence systems. The development of legal norms and standards that will ensure the transparency of algorithmic models and minimize the risks of bias is relevant. In addition, research prospects are related to solving the problems of interdisciplinary cooperation to ensure the most*

*complete understanding of the social, legal and ethical consequences of the use of algorithms in decision-making.*

***Key words:*** *artificial intelligence, decision-making, algorithms, transparency, controllability, fairness, ethics, legal regulation.*

**Problem statement**. Today, the use of AI in critical areas - from medical diagnoses to court decisions - is not just a trend, but a reality that poses a number of profound ethical challenges to society. On the one hand, powerful algorithms improve efficiency and accuracy in many industries: for example, about 43% of healthcare leaders are already using AI to monitor patients in hospitals, and by the end of 2025, about 90% of hospitals plan to implement AI systems for diagnosis and monitoring. On the other hand, these algorithms pose significant risks:

- biases and discrimination: AI often transfers biases from training data into its decisions. For example, in face processing systems for women with dark skin, the error rate reached 34,7 %, while for light-skinned men it was only 0.8%. An analysis of criminal court systems showed that algorithmic systems were 45% more likely to impose harsher sentences on people of African American descent.
- opacity and "black boxes": high-performance neural networks are opaque - even developers often cannot explain the solution mechanism. Only 35% of consumers fully trust AI solutions;
- lack of control: systems act autonomously without clear designation of responsible persons and reinforces the phenomenon of "moral neglect" when responsibility is transferred to the machine;
- automation versus human control: there is a hypothesis that increasing the autonomy of AI without proper supervision will contribute not only to mistakes but also to "departure" from human values in decisions;

- decrease in trust and rejection: the phenomenon of algorithm aversion - disdain for algorithmic solutions - is manifested in many areas: healthcare, employment, law;
- the problem of "deepfake" and disinformation: AI is capable of generating fake content that undermines public trust in information sources and poses a threat of manipulation;
- environmental impact: training large models requires significant energy resources, which poses an additional ethical challenge.

Thus, today there is an urgent need for an interdisciplinary approach - external audits, regulation of autonomy, and institutionalization of ethical responsibility. Without these components, AI in decision-making can lead to serious social, legal, and environmental consequences.

**Analysis of recent research and publications.** A study of the current research and publications in the area of ethical challenges of artificial intelligence shows an increasing focus on the dimensions of fairness, transparency, and responsibility of algorithmic systems. In their article "Big data's disparate impact", Barokas and Zelbst [1] examine the way in which big data can bring discrimination and biases against some social groups, by demonstrating legislative as well as technological measures and precautions that such phenomena require. Their importance lies in structural biases that result from the data quality and processing methods.

Binns [2] considers the matter of fairness in predictive algorithms in the context of political philosophy, providing a rich area of concepts to consider how algorithms might exacerbate or reduce social inequality. This way we can better grasps the ethical dimension of automated decisions and we are able to ground ethical principles.

Danks and London [3] address the issue of algorithmic bias in autonomous systems, and argue the importance of developing standard methods for detecting and mitigating such biases. Their work offers usable critiques for

the testing and development of algorithms from security and ethical perspectives.

Doshi-Velez and Kim [4] further the pursuit of interpretability of machine learning by introducing the concept of a scientific basis for explainability of AI models. Their work is important for increasing transparency and user confidence in algorithmic solutions.

Mittelstat [5] systematizes ethical issues related to algorithms and outlines the main areas of scientific discussion, including privacy, responsibility, bias, and transparency. Their analysis creates a framework for further study and development of ethical practices in the field of artificial intelligence.

At the same time, the problematic issues of ethical challenges in the application of artificial intelligence in decision-making based on the coordination of the interests of all stakeholders and the combination of different approaches, processes, technologies, resources in order to achieve the overall goals of artificial intelligence development remain unresolved.

**The purpose of the article** is to determine the ethical aspects of the use of artificial intelligence in decision-making and to develop practically oriented recommendations designed to ensure transparency, controllability and a socially acceptable level of use of algorithms.

**Summary of the main material.** In today's world, artificial intelligence is increasingly a part of decision-making in a wide array of fields, from medicine and the law to advertising and hiring. But the use of algorithms will clearly raise all sorts of ethical issues that will need to be carefully studied and regulated. Some of the most serious issues related to bias in algorithms are caused by unbalanced, ill-annotated training data, or mis-representation of social biases in the model. Bias introduces discrimination in respect to different people, where the fairness of decisions is in doubt. The second significant problem is the dearth of openness and explainability, the black box nature of algorithmic decision-making. The authors of the analysis point out that it's

difficult to understand even for the developers why a decision was reached with some neural networks. This fact is hard to control and gives rise to lack of trust from users and society at large. Absence of transparency makes it difficult to judge whether the intention is to ensure decisions are in accordance with ethical standards and human rights.

Yet another serious challenge is who controls and is held accountable for decisions taken with the assistance of AI. As autonomous systems perform themselves it is quite common to ask: who is (or would be) responsible if something goes wrong and there are negative consequences - developers, operators, or the algorithms, what adds complexity in terms of legal regulation and protection the rights of the affected ones.

An additional issue is how to balance between decision making automation and human decision making. Automation is an efficiency tool, but the complete transferring of responsibility to machines can generate alienation from human values, ethical considerations and the context of situations. There is the need to support mechanisms, which enable humans to control and correct decisions of artificial intelligence [6]. The privacy and data security implications of the algorithms cannot be dismissed.

Training models requires a lot of sensitive personal and private information, leading to potential data leakage and misuse. A deficit in general regulation and appropriate standard of data protection can result in abuse of privacy and in the growth of social inequality.

In general, ascertaining such challenges and considering them systematically is necessary to develop computationally reliable, fair and socially acceptable algorithmic systems that can make society better off without causing damage to specific groups or individuals. Issues of principle on artificial intelligence, such as algorithmic bias, lack of transparency, and control, are driven by a number of factors. Above all else, the bias of AI algorithms is often

a product of the quality and make-up of the data on which the models are trained.

If the data reflects the preexisting biases, or long-held stereotypes, or social disparities either in an official or unofficial way, or doesn't have sufficient information regarding given groups, the algorithm will by default reproduce and sometimes reinforce those very biases.

In this framework, these models act as "black boxes" where thousands of parameters and weights are applied to each other, and it becomes impossible to explain why a decision was taken in a simple way. This lack of transparency makes auditing algorithms difficult and does not allow catch any errors or ethical overreach. Thus, users and regulators are never able to fully trust the output of the systems, and developers are usually unable to effectively watch or tune how the models behave. The lack of being in control is for technical, technical, organizational and legal reasons.

Firstly, in the case of autonomous algorithms deciding independently without constant human intervention, this poses a risk of uncertainty in liability. On the other hand, there is no clear set of norms and standards regarding AI and the circle of responsible actors.

The absence of clear and open audit and control trails would mean that in an error or damage case, it would not be possible to unequivocally determine who should be held responsible (developer, user, or even the system) [7].

Another way in which these problems can appear is through institutional unpreparedness and the absence of an ethical culture in the AI development and use. Too many businesses focus on what is technologically possible and economically advantageous while giving insufficient attention to ethical considerations and hazards. This results in disregard for bias testing procedures, lack of transparency in communication with users, and no feedback loops.

The issue of suboptimal regulation is, however, compounded by loss of sovereignty. In the majority of countries, legislation lags behind the

technological progress and the so called "gaps" appear in the regulation. This results in legal uncertainty and problems of enforcement following breach or damage generated by algorithmic solutions.

Therefore, the roots of ethical issues in AI usage are multifaceted and multinodal, ranging from objective technical constraints or data quality to organizational, cultural, and regulatory aspects. Understanding these sources is crucial in order to design effective means for counteracting bias, achieving transparency and controllability in algorithmic systems.

Practical principles should be designed and enforced in society, laws, and technology to guarantee the ethical usage of AI techniques in decision-making.

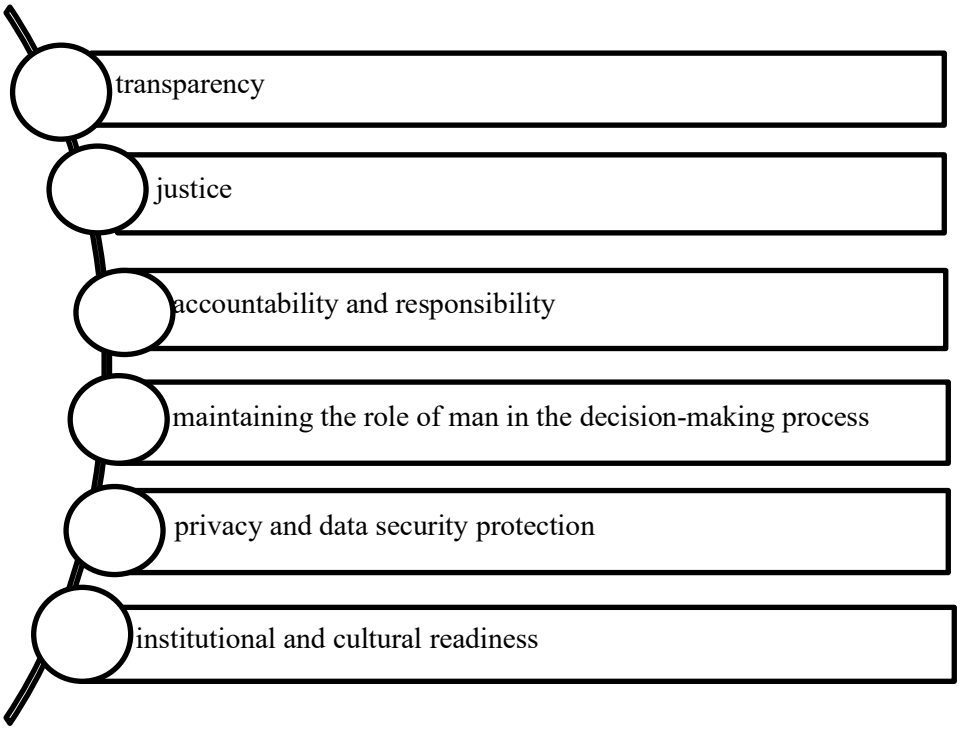


**Fig. 1. Practically oriented principles to ensure the ethical application of artificial intelligence algorithms**

First and foremost it is crucial to adopt the principle of transparency, that means making algorithmic processes accessible and understandable for end users and regulators, and this can be achieved with the support of Explainable AI technologies that enable to explain the logic of decision making and to reduce "black box" in the functioning of models.

The other central principle is fairness – we cannot have the public using systems that produce biased results and we need to clear the market of biased models. To do this, there should be regular audits of data and model for gender, age, race, or social status bias. Developers should build systems that enable the revision of the algorithms in a way that would mitigate these biases so that all users have equal opportunities.

Third is the principle of accountability and responsibility. Training clear protocols that would set out who is responsible for building, deploying and overseeing algorithmic systems would allow for the quick identification and addressing of mistakes. Another key challenge is to devise regulatory mechanisms that would establish legal responsibility in human rights infringements as a result of AI errors or biases [8]. The fourth principle is to preserve the role of humans in the decision-making process (human-in-the-loop). However, even with a high level of automation, there should be a place for human intervention in a critical case, where a wrong decision may have implications. This approach will increase trust in the systems and allow correcting the errors of a potentially "flawed" algorithm. Secondly, it should be based on the principle of the privacy and security of data. The confidentiality of personal data and modern encryption technologies, mandatory for handling large volumes of information, are essential. Data collection minimizate and data process transparency policies contribute to the ethical application of AI. Finally, organizations' institutional and cultural readiness for the responsible use of AI is vital. It implies training, awareness of ethical risks, and the adoption of corporate codes of ethics. This holistic multi-pronged approach will minimize the risks associated with the application of the technology and build trust in it. In order for algorithmic systems to be fair and controllable, it is necessary to ensure the interaction of legal, social, and technical toolsets that, jointly, create an effective control mechanism. From a legal standpoint, it is legislative regulation. Laws should stipulate clear rules for the development, testing, and

use of AI algorithms. They must also contain the requirements for transparency, human rights protection, non-discrimination, and the development of measures for violators. The implementation of the stated norms should stimulate developers to be guided by ethical standards, to provide maximum protection for the user. The social toolset includes the formation of an informed and ethical society regarding AI. The heat of educational programs for both specialists and the general public problematic situations and benefits of such systems. Public control, the participation of stakeholders, and an open dialogue of developers, users, and regulatory authorities are designed to ensure the transparency and fairness of the implementation of technology. Technical tool is the basis for accountable and fairness. These include auditing and auditing methods to identify bias, errors and "inappropriate" behavior of models. Models can be understandable with the help of Explainable AI, and it is possible to be trainable in such a way that the information "leaks" cannot occur have been instances have such monstrosity.

In addition, technologies to protect data protection, such as differential privacy and encryption methods, are also needed in order to guarantee both information security and; minimal leaks. Moreover, the technical tool is the effective implementation of human-in-the-loop when a human is responsible for the final decision in critical cases to allow adjusting the algorithm's results, to avoid errors, and to ensure ethical decisions. Laws will also increase user trust. Also, the development of standards and certification mechanisms that ensure the compliance of the application with ethical and data security standards, will also be of encompassing importance. It may include bias verification, legal compliance, or control mechanism availability; they definitively add a layer of protection; it is evident that all of the above will contribute to the quality of the implementation of the technology.

**Conclusions and Prospects for Further Research**. Thus, the study has shown that the implementation AI algorithms in decision-making and decision-

support systems goes with various 1 considerable ethical problems: algorithmic biases, lack of transparency and clear accountability, complex biases that are connected both with the quality of the data and the properties of the models, and weak legal and organizational control.

The analysis showed that to minimize the negative consequences, it is necessary to implement the principles of transparency, fairness, responsibility, and preserve the role of the individual in the decision-making process.

Practically oriented recommendations include the development and application of Explainable AI technologies, regular bias audits, clear identification of responsible persons, and personal data protection. It is important to combine technical, social, and legal measures to create a comprehensive system of control and ensure the ethics of algorithmic systems.

## References

1. Barocas, S., & Selbst, A. D. (2016). Big data's disparate impact. *California Law Review, 104*(3), 671–732. https://doi.org/10.2139/ssrn.2477899

2. Binns, R. (2018). Fairness in machine learning: Lessons from political philosophy. *Proceedings of the 2018 Conference on Fairness, Accountability, and Transparency*, 149–159. https://doi.org/10.1145/3287560.3287598

3. Danks, D., & London, A. J. (2017). Algorithmic bias in autonomous systems. *Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence (IJCAI-17)*, 4691–4697. https://doi.org/10.24963/ijcai.2017/654

4. Doshi-Velez, F., & Kim, B. (2017). Towards a rigorous science of interpretable machine learning. *arXiv preprint arXiv:1702.08608*. https://arxiv.org/abs/1702.08608

5. Mittelstadt, B. D., Allo, P., Taddeo, M., Wachter, S., & Floridi, L. (2016). The ethics of algorithms: Mapping the debate. *Big Data & Society, 3*(2), 1–21. https://doi.org/10.1177/2053951716679679

6. Selbst, A. D., & Barocas, S. (2018). The intuitive appeal of explainable machines. *Fordham Law Review, 87*(3), 1085–1139.

7. Shneiderman, B. (2020). Human-centered artificial intelligence: Reliable, safe & trustworthy. *International Journal of Human–Computer Interaction, 36*(6), 495–504. https://doi.org/10.1080/10447318.2020.1741118

8. Wachter, S., Mittelstadt, B., & Floridi, L. (2017). Transparent, explainable, and accountable AI for robotics. *Science Robotics, 2*(6), eaan6080. https://doi.org/10.1126/scirobotics.aan6080

9. Whittaker, M., Alper, M., Bennett, C., Hendren, S., Kaziunas, E., Mills, M., Gebru, T. (2018). AI now report 2018. AI Now Institute, New York University. https://ainowinstitute.org/AI_Now_2018_Report.pdf

10. Zarsky, T. Z. (2016). The trouble with algorithmic decisions: An analytic road map to examine efficiency and fairness in automated and opaque decision making. *Science, Technology, & Human Values, 41*(1), 118–132. https://doi.org/10.1177/0162243915605575