ARTIFICIAL INTELLIGENCE AND HUMAN RIGHTS: PHILOSOPHICAL AND ETHICAL FOUNDATIONS OF LEGAL REGULATION

Popovych Terezia

DOI: https://doi.org/10.61345/1339-7915.2025.2.18

Annotation. The article examines the philosophical and ethical foundations of legal regulation of artificial intelligence in the context of protecting fundamental human rights. The author analyzes the transformation of traditional ethical concepts in the conditions of rapid development of AI technologies and their integration into critically important spheres of social life. Special attention is paid to the problem of balancing the innovative potential of artificial intelligence with the need to protect human rights and freedoms. The concepts of "artificial will" as a mechanism for embedding ethical principles into AI algorithms are considered, as well as challenges related to the limitations of traditional means of state coercion in the context of the global nature of AI technologies. The necessity of creating international mechanisms of control and coordination in the field of ethical regulation of artificial intelligence is substantiated. The author proves that effective AI regulation should be based on a human-centered approach, where the motive is not the restriction of technological progress, but the protection of humans from potential negative consequences of using AI systems. The research contributes to the development of theoretical and legal foundations for ethical regulation of artificial intelligence and defines prospects for forming a new paradigm of legal regulation in the digital age.

Key words: artificial intelligence, human rights, ethics, legal regulation, algorithmic justice, digital ethics, philosophy of law

1. Problem Statement

The rapid development of artificial intelligence technologies in the 21st century is fundamentally changing the social, economic, and legal realities of modern society. AI systems are increasingly integrating into critically important spheres of human life – from medical diagnostics and judiciary to education and social welfare. However, this technological progress generates a number of fundamental ethical dilemmas and legal challenges that require urgent resolution.

The central problem is the conflict between AI technological capabilities and the need to protect fundamental human rights. Algorithmic decisions can violate principles of justice, equality, and non-discrimination, especially when AI systems demonstrate bias toward certain social groups. At the same time, automation of decision-making processes calls into question the human right to fair trial and the possibility of appealing algorithmic conclusions.

Particularly acute is the problem of the absence of clear legal frameworks and ethical standards for the development, implementation, and use of AI technologies. This creates legal uncertainty and makes effective protection of citizens' rights in the digital age impossible. Moreover, the transnational nature of AI technologies requires coordination of international efforts to create universal ethical principles and legal mechanisms.

Thus, the relevance of the research is determined by the need to find an optimal balance between the innovative potential of artificial intelligence and the inviolability of fundamental human rights, which is a key condition for sustainable development of digital society.

2. Research Objective

The objective of the research is to form theoretical and legal foundations for ethical regulation of artificial intelligence through the prism of human rights protection, develop conceptual approaches to balancing the innovative potential of AI technologies with the inviolability of fundamental human rights, and determine optimal mechanisms of legal regulation in the field of artificial intelligence taking into account philosophical and ethical principles and international experience.

3. Analysis of Source Base

It should be noted that the source base for researching the problem of artificial intelligence, ethics, and human rights is characterized by interdisciplinarity and rapid development, reflecting the novelty and relevance of the topic. Literature analysis allows identifying several main groups of sources:

a) Normative-legal sources constitute the fundamental basis of the research. These include international documents, particularly "UNESCO Recommendations on the Ethics of Artificial Intelligence" (2021), "Ethics Guidelines for Trustworthy AI" by the European Commission (2019), as well as the draft EU Regulation on Artificial Intelligence (AI Act, 2024). National legislations of various countries, including laws on personal data protection and algorithmic transparency, also constitute an important part of the legal base;

b) Scientific monographs and collective works are represented by works of leading researchers in the field. Of particular value are the works of L. Floridi on digital ethics, C. O'Neil's "Weapons of Math Destruction" on algorithmic bias, as well as collective monographs edited by R. Calo and M. Dąbrowski on AI and human rights issues;

c) Periodical publications include articles in leading international journals: "AI & Society", "Ethics and Information Technology", "Computer Law & Security Review", "Information, Communication & Society". Domestic publications are represented in journals "Law of Ukraine", "Information and Law", "Scientific Notes of NaUKMA";

d) Analytical reports of international organizations constitute a separate group of sources. These are reports from the UN, OECD, World Economic Forum, as well as specialized organizations such as Partnership on AI, Future of Humanity Institute, AI Now Institute.

At the same time, it should be noted that current limitations of the source base are related to the speed of technological development, which leads to rapid obsolescence of information. Moreover, a significant part of research is conducted in the English-speaking academic environment, which creates certain geographical and linguistic limitations in covering diverse cultural and legal approaches.

4. Research Material Presentation

In a democratic rule-of-law state, the individual constitutes the highest social value, playing a key role in all spheres of life – material, political, social, spiritual-cultural. At the same time, each person is individual, unique, unrepeatable, singular. Their involvement in the activities of a rule-of-law state is determined by material and spiritual capabilities regarding the use of their own subjective rights and voluntary fulfillment of legal obligations. The legal status of a person consists of rights, freedoms, and obligations defined by legislation, which constitute possibilities that are realized through their practical implementation [1, p. 933].

The necessary stability of the concept of protecting human and citizen rights and freedoms is achieved through reliance on a system of principles tested by science and practice. The viability and progressiveness of this concept is formed through the combination of legal, moral, traditional, and other socio-regulatory norms. Taking such an approach into account will prevent legal negativity

from becoming a dominant phenomenon of the legal system and will restrain the pressure of legal nihilism and indifference [2, p. 449].

Ethics constitutes the foundation for forming the legal culture of the state and society as a whole. The legal culture of European states began intensive development from the moment of adopting the French Declaration of the Rights of Man and of the Citizen. This act declared human freedom, but only within limits that do not violate the freedom of other persons. Such an approach first defined relatively clear boundaries of ethical foundations of freedom and human rights. Thus, not merely moral guidelines based on abstract spiritual and philosophical concepts were established, but specific boundaries of human freedom were defined and the principle of equality was enshrined. From this period, foundations of equality among society members, respect for private life, and mutual respect began to be established in society. Subsequently, these principles were embodied in legislation and other normative acts regulating social relations.

Ethics is the science of human attitude toward morality itself: about what meaning, what internal necessity a person sees in accepting certain moral norms, on what they base their choice, where the need for moral self-restraint arises in them at all. Without addressing this sphere of theoretical and practical universals (the greatest generalizations describing a certain state of the world and human worldview in general), we will not be able not only to clarify the actual moral quality of a particular phenomenon, but even to determine its fundamental relation to the field of morality as a whole. However, these universals are the subject of reflections and searches of philosophy as a special branch of human cognition. Hence it follows that ethics, at least in its essential foundation, can only be a philosophical science (which, of course, does not exclude the existence of partial branches of ethical cognition distant from philosophical themes – say, various applied professional ethics, empirical moral studies, etc.) [3, p. 5].

I. Gizel and other researchers adhered to the position according to which endowing a person with the capacity for free choice simultaneously places on them the obligation to bear responsibility for their own actions, primarily before themselves, as well as before society [4, p. 153].

Undoubtedly, such an approach indeed permeates the modern foundations of legislation. However, applying similar criteria and beliefs to regulating human-AI interaction may generate certain difficulties, as in the modern world, a person using various AI systems is deprived of the opportunity to consciously choose and understand the essence of this technology and potential consequences of its application. This example illustrates the importance and urgency of forming a transparent environment for AI development that will be based on human interests and rights [5].

Another philosophical-ethical aspect directly related to the subject of research is the problem of the relationship between human will and intellect. The above-cited scholar noted that will plays a leading and determining role in human life compared to intellect, as it is based on virtue. At the same time, the researcher considered the opposite position, asserting that intellect is a more reliable criterion, as it cognizes the environment through comparison and analysis, while will is based on perception of external signs. The scientist emphasized that humans possess freedom precisely due to their capacity for intellectual activity [4, p. 154].

Projecting these reflections onto modern human-AI relationships, one can note the relevance of the researcher's position. First of all, currently the level of human intelligence, particularly the ability to understand one's own rights and analyze the possible impact of innovative technologies on their realization, becomes crucial for the future coexistence of these seemingly separate worlds – real and virtual. At the same time, the desire for rapid development and new prospects opened by AI technologies may overshadow other needs and hinder deep analysis and research of innovative technologies. In this aspect, studies on the influence of will on our desires and motivations also deserve attention. Therefore, the researcher emphasized that people often act under the influence of emotions without restraining themselves, and it is precisely in such situations that will functions as a filter capable of directing human thinking and behavior in a constructive direction [4, pp. 154-155].

From this perspective, the question of implementing so-called "artificial will" in AI technologies attracts attention. Probably, precisely embedded algorithms containing a complex of fundamental

principles of justice, humanity, ethics, equality, will be able to serve as that protective mechanism against the negative impact of such technologies on human life and rights. After all, similar to regulating interpersonal relations in society, when violations of some people's rights by others occur or in the absence of proper legal regulation or ineffectiveness of human rights protection norms, the only mechanism and guideline are always fundamental principles and ethical beliefs. There are grounds to assume that in the field of technology, such an approach also has prospects for implementation. After all, when law loses effectiveness under normal life conditions or in case of violation of generally recognized conditions of life activity, such as during war, the only factor capable of maintaining the proper course of life is ethical and moral foundations of existence [6, p. 3].

On the other hand, law should always be based on moral norms. This is an ideal scenario when legal norms fully harmonize and do not conflict with moral norms. Under such conditions, law is effective and capable of regulating social relations in the long term. This constitutes the foundation of a rule-of-law state. However, since we exist in an imperfect society, the main means of law implementation remains coercion. The function of coercion is assigned to the state. However, for coercion to be applied to the extent that it will perform the function of law implementation rather than becoming an instrument of unlawful restriction of human rights, it must also be legitimized by the moral foundations of society. In the science of ethics, coercion is possible exclusively in relationships between people, as they are free to choose their behavior, analyzing the consequences of their own actions, and consciously agree to this. In other words, a person, violating legal norms and established rules, consciously agrees to the application of "violence" in the form of coercion to certain behavior. If we apply the same approaches developed by science over a long time to relations related to artificial intelligence, this will cause chaos, at least in the initial stages of regulating these relations. Chaos, in turn, is interpreted by philosophers as a transition to something new, so probably this is the only possible way to achieve a new order of social life in which the role of humans may transform in connection with the emergence of a new "subject" – AI. This will require, in turn, developing new means of coercion that will serve both as protective mechanisms for human rights and restrictive ones – regarding unlawful actions of AI technologies [7, p. 390].

Returning to the problem of state monopolization of coercion, one can identify a number of issues that are currently not the subject of scientific research but may require the development of new regulatory instruments in the future. For example, given the global nature of corporations engaged in AI development and application, and their lack of attachment to a specific state, the ability to apply coercive measures to such participants by the state in the format in which this currently occurs is limited. In this regard, the expediency of reviewing approaches to monopolizing coercion in state hands and implementing international control mechanisms in this field is being considered. Similar approaches have already begun to be implemented by individual states in the field of regulating the tax activities of such AI-developing corporations [7, p. 391].

In the context of means of regulating social relations, it is also appropriate to pay attention to the purpose of such regulation. Considering this issue through an ethical prism, most researchers have always analyzed the choice of means for regulating certain relations in view of the goal to be achieved. At the same time, for example, W. Wundt argues that there are discrepancies and factors between the means of achieving the goal and the goal itself that humans cannot always influence [8, p. 33].

This idea was considered by numerous philosophers who studied issues of ethics, morality, good and evil. Indeed, there are many contradictions in the world, and the process between actions and consequences is not always linear and predictable. For example, good deeds often lead to negative results, while negative and not entirely moral actions can cause positive consequences. Applying this approach to human-AI relations, it seems that precisely such a development scenario may occur. Namely, when the main motive of AI developers and users is achieving economic benefit and influence, the overall impact on social relations from the development of such technologies may turn out to be positive and cause significant changes both in people's lives and in the world as a whole. Today this is already manifested in achievements in the medical field that save lives, in the

security sphere when AI technologies help prevent irreversible consequences of war, and in many other spheres of our existence [9].

Interesting in this context is also the attitude of Machiavellians toward ethical principles of activity in society, which actually has a huge number of supporters. They argued that the problem of the relationship between means and ends, which has been discussed by researchers for centuries, is artificial, as they believed that a "high" goal justifies any means. As we observe from practical life, some companies or individuals, gaining power and influence, often act precisely according to such rules, achieving their goals by any means and even ignoring moral principles and law. That is why we are convinced that the approach once proposed by N. Machiavelli is unacceptable for relations concerning human rights. At the same time, for legal regulation to be not only theoretical but also practically applicable, we should not cherish illusions that creating regulation based on moral and ethical norms will immediately become a panacea for solving all problems. That is why when developing mechanisms for regulating these relations, it is necessary to take into account the interests of those who act in this field and plan everything so that there is no need and desire on the part of those who should "submit" to these rules to violate them or seek workarounds, as excessive formalism and regulation by the state is currently a negative factor of the national legal system [6, p. 8].

Through the prism of philosophical-ethical theories, it is also appropriate to mention the question of motive. The main motive for regulating AI technologies should be not punishment for rapid and uncontrolled development of such technologies, but protection of humans from possible negative impact of such technologies on their life and rights. That is why, when developing moral foundations for the functioning of artificially created systems, one should be guided by the motive of creating an environment favorable for the development of such technologies while simultaneously respecting fundamental human rights, which are priority over technological achievements and technological progress as such. Having such a motive and goal, responsible persons will, accordingly, choose adequate means to achieve it [7, p. 390].

5. Conclusions

The conducted research on philosophical-ethical foundations of artificial intelligence regulation in the context of human rights protection allows formulating the following conclusions:

First, the modern development of artificial intelligence technologies creates fundamentally new challenges for traditional concepts of human rights and legal regulation. The application of classical ethical principles based on freedom of choice and human responsibility to the AI sphere reveals its limitations, as users of AI systems are often deprived of the opportunity to fully understand the nature of technology and consequences of its use. This requires the formation of new approaches to ethical regulation that will combine transparency of technological processes with ensuring human rights and interests.

Second, effective legal regulation of artificial intelligence requires rethinking traditional mechanisms of coercion and control. The global nature of AI corporations and transnationality of technologies limit the possibilities of national states to apply traditional means of legal coercion. This actualizes the need for developing international control and coordination mechanisms, as well as implementing "artificial will" – ethical algorithms containing fundamental principles of justice, humanity, and equality as internal regulators of AI systems.

Third, successful regulation of artificial intelligence should be based on a correctly defined motive – not on punishment for technological progress, but on protecting humans from potential negative consequences. This involves creating an innovation-friendly environment while simultaneously ensuring the priority of fundamental human rights over technological achievements. Practical implementation of such an approach requires taking into account the interests of all process participants and forming regulatory mechanisms that will minimize incentives for violating ethical norms and circumventing legal restrictions.

References:

1. Bielova, M.V., Byelov, D.M. (2024). Humanistychni zasady pravovoi systemy: dialektyka teorii ta praktyky v umovakh suchasnykh vyklykiv [Humanistic principles of the legal system: dialectics of theory and practice in the conditions of modern challenges]. *Analitychno-porivnialne pravoznavstvo – Analytical and Comparative Jurisprudence*, 5, 933-939 [in Ukrainian].
2. Bielova, M.V., Byelov, D.M. (2023). Implementatsiia shtuchnoho intelektu v dosudove rozsliduvannia kryminalnykh sprav: mizhnarodnyi dosvid [Implementation of artificial intelligence in pre-trial investigation of criminal cases: international experience]. *Analitychno-porivnialne pravoznavstvo – Analytical and Comparative Jurisprudence*, 2, 448-454 [in Ukrainian].
3. Etyka: Navchalnyi posibnyk [Ethics: Textbook]. (2005). Kyiv: Vydavnychyi tsentr «Akademiia» [in Ukrainian].
4. Gizel, I. (2013). Tvir pro vsiu filosofiiu [Work on all philosophy]. *Sententiae*, 2, 152-162 [in Ukrainian].
5. Rodrigues, R. (2020). Legal and human rights issues of AI: Gaps, challenges and vulnerabilities. *Journal of Responsible Technology*, 4 [in English].
6. Roumate, F. (2021). Artificial Intelligence, Ethics and International Human Rights Law. *The International Review of Information Ethics*, 29, 1-10 [in English].
7. Jobin, A., Ienca, M., & Vayena, E. (2019). The global landscape of AI ethics guidelines. *Nature Machine Intelligence*, 1, 389-399 [in English].
8. Vundt, V. (1992). Systema filosofii [System of philosophy]. Kyiv [in Ukrainian].
9. Kazim, E., & Koshiyama, A.S. (2021). A high-level overview of AI ethics. *Patterns*, 2(9), 100314 [in English].

Terezia Popovych
Candidate of Law, Associate Professor,
Associate Professor of the Department of Theory and History of State and Law
State University of Ukraine "Uzhhorod National University"
ORCID: 0000-0002-8333-3921