THE BALANCE BETWEEN INNOVATION AND HUMAN RIGHTS: PROBLEMS OF APPLYING ARTIFICIAL INTELLIGENCE

Popovich Terezia

DOI: https://doi.org/10.61345/1339-7915.2025.3.16

Annotation. The article provides a comprehensive analysis of the impact of artificial intelligence technologies on the human rights system in the context of the digital transformation of society. The main threats and challenges posed by AI systems for the implementation of fundamental human rights and freedoms are investigated. Particular attention is paid to the problems of algorithmic bias, which leads to a violation of the right to non-discrimination, mass collection and processing of personal data without proper control, which threatens the right to privacy, restrictions on freedom of expression through automated content moderation, as well as threats to the right to a fair trial in the case of using automated decision-making systems without proper transparency.

The specific risks associated with mass surveillance and biometric identification technologies, including facial recognition systems in public places, the use of AI in employment and the military, and the manipulation of public opinion through deepfake technologies, are analyzed. Three stages of assessing the impact of AI on human rights are considered: analysis of the quality of training data, risk assessment at the system design stage, and consideration of algorithmic interactions. It is argued that AI systems, by their nature, reproduce social biases embedded in past experience data and do not have the inherent ability to change their behavior in accordance with the evolution of ethical norms in society. The application of artificial intelligence in the financial sector is examined in detail, in particular in credit scoring systems, where algorithms analyze huge amounts of data about the applicant's digital footprint. The problem of "network discrimination" is identified, when a person's financial capabilities are assessed based on the characteristics of their social environment, which violates the principle of individual responsibility and can limit freedom of belief through self-censorship. The example of the practice of American companies shows how the use of AI systems in financial decision-making can both expand access to credit for representatives of marginalized communities and strengthen existing forms of discrimination.

Key words: artificial intelligence, human rights, algorithmic discrimination, right to privacy, digital technologies.

1. Statement of the problem.

The current stage of technological development is characterized by the rapid penetration of artificial intelligence into all spheres of human life – from the everyday use of personal assistants to critical industries such as justice, finance, healthcare and national security. AI systems are increasingly making or significantly influencing decisions that directly affect human rights and freedoms: determining access to credit, assessing the risks of recidivism in criminal justice, selecting job candidates, moderating content in the online space and even being used in military operations. At the same time, along with the undeniable advantages that artificial intelligence technologies offer to society - increasing efficiency, speed of information processing, automation of routine processes – fundamental questions arise regarding their impact on the human rights system.

Of particular concern is the fact that the development and implementation of AI systems often outpaces the formation of an adequate legal framework and ethical standards for their use. Algorithmic bias, mass collection of personal data, opacity of decision-making processes, risks of discrimination and restrictions on fundamental freedoms create a threat of "erosion" of human rights in the digital age.

The lack of proper control over the functioning of artificial intelligence systems, insufficient public awareness of the mechanisms of their operation and limited possibilities of appealing automated decisions make it necessary to conduct a comprehensive study of the impact of AI on human rights. This issue is of particular importance in the context of Ukraine's European integration aspirations and the need to bring national legislation into line with European standards for the protection of human rights in the context of the digital transformation of society.

2. Analysis of the source base.

The source base of the study is a set of analytical documents, scientific publications and regulatory legal acts that consider the issue of the impact of artificial intelligence on human rights from various methodological positions. The fundamental source is the analytical report of the Verkhovna Rada of Ukraine Commissioner for Human Rights "Human Rights in the Age of Artificial Intelligence" [1], which systematizes the main threats to human rights caused by the use of AI technologies, in particular the violation of the right to non-discrimination due to algorithmic bias, the right to privacy, freedom of expression and the right to a fair trial. This document also highlights the risks associated with mass surveillance technologies, biometric identification and autonomous weapons systems. An important contribution to understanding the theoretical aspects of the problem was made by the studies of R. van Est, J. Gerritsen and L. Kuhl [3], who distinguish three conceptual approaches to the interaction of AI and human rights: the use of AI as a factor of threat to human rights, as a tool for improving the quality of their protection, and considering AI as a potential subject of rights. H.-Ya. Liu and K. Zaviska [6] summarize the main risks to ensuring proper protection of human rights, emphasizing the problem of fragmented consideration of relevant issues and the need for a comprehensive approach to regulating the use of AI technologies.

Sectoral aspects of the application of artificial intelligence are considered in the works of Kravchuk S.M. [2], which analyzes general recommendations for the sustainable implementation of AI taking into account the protection of human rights, and Kozmenkova M.G. [11], which explores the opportunities and challenges of using artificial intelligence in financial institutions in Ukraine, in particular in the context of the development of the ecosystem of online services. The study pays special attention to the problems of using AI in the field of lending, where the widespread use of algorithms to assess the creditworthiness of clients creates risks of "network discrimination" and restrictions on freedom of belief [12; 13; 14].

The methodological basis is formed by works that consider the stages of assessing the impact of AI on human rights [8], including the analysis of the quality of training data, risks at the stage of system design and taking into account algorithmic interactions. Separate studies [9; 10] are devoted to the problems of unethical data reproduction by AI systems and conflicts between the needs of AI in data sets and the right to privacy. A comprehensive analysis of the source base confirms the interdisciplinary nature of the issue of the impact of artificial intelligence on human rights, which requires the integration of legal, technical, ethical and sociological approaches to its solution.

3. Purpose of the study.

The purpose of the study is to analyze individual aspects of the impact of artificial intelligence technologies on the human rights system, identify the main threats and challenges that AI systems create for the implementation of fundamental rights and freedoms, as well as develop recommendations for the formation of legal and ethical mechanisms for ensuring the protection of human rights in the conditions of the mass introduction of intellectual systems into various spheres of public life.

4. Presentation of the research material.

Today, artificial intelligence is used in almost all areas of human life. The latest technologies are able to perform a variety of tasks: driving a car, production processes at enterprises, generating text, music, recognizing people's faces and voices, performing the functions of a personal assistant

in a smartphone, etc. They are built into various devices that are used daily in public policy, urban infrastructure, business or simply in everyday life. Despite the broad prospects for society that intelligent systems can create, it is also necessary to pay attention to the ethical and legal aspects of their use, in particular the impact on human rights and freedoms [1].

So, as we can see, the informatization of society and the introduction of information and communication technologies have become an integral part of our lives. They provide quick access to information, simplify communication and open up new horizons for learning and work. However, the growth of data volumes and their processing using AI raise concerns about the protection of human rights [2, p. 101].

R. van Est, J. Gerritsen and L. Kuhl believe that developments in the field of AI are almost completely undeveloped, despite the significant impact of these technologies on human rights, and see this as a threat to the erosion of human rights. In the context of ensuring the protection of human rights, taking into account the widespread use of AI technologies, scientists distinguish three approaches: the use of AI as a factor that threatens the realization of human rights; the use of AI to improve the quality of ensuring the protection and observance of human rights; AI as a subject that can and should be endowed with "human rights" [3]. It is often noted that the widespread use of AI can contribute not only to the emergence of some local problems of a legal and practical nature in specific individual areas, but can also threaten the general interests of man.

The Commissioner for Human Rights of the Verkhovna Rada of Ukraine in his analytical report identifies the following main threats: violation of the right to non-discrimination due to algorithmic bias, when AI systems reproduce or reinforce existing social stereotypes regarding race, gender, age or other characteristics; violation of the right to privacy due to the mass collection, processing and analysis of personal data without proper control; restriction of the right to freedom of expression through automated content moderation and censorship in the online space; threats to the right to a fair trial in the case of the use of automated decision-making systems in justice without proper transparency and the possibility of appeal [1].

Of particular concern are technologies of mass surveillance and biometric identification, in particular facial recognition systems in public places, which can be used for total control over the movement of citizens and the suppression of freedom of assembly and peaceful protests. In addition, the use of AI in the field of employment (automated selection of candidates, monitoring of employee productivity) can lead to discrimination and violation of labor rights. In the military sphere, the development of autonomous weapons systems (lethal autonomous weapons) threatens compliance with international humanitarian law and the right to life [1]. In addition, it should be noted the risks of manipulating public opinion through targeted disinformation and deepfake technologies, which undermines democratic processes and the right to access reliable information [5, p. 108]. In turn, H.-Ya. Liu and K. Zaviska summarize the following main risks for ensuring proper protection of human rights in view of the challenges posed by the use and development of AI technologies, namely: considering relevant problems separately depending on the sphere of activity or the nature of the right that may be violated, and not in a complex; building modern approaches to the problem of ensuring the protection of human rights, mainly regarding the issue of interaction between a person and the state [6].

It should be noted that today the following assessment of the impact of AI on human rights is also distinguished:

– assessment of the quality of training data. At this stage, it is necessary to analyze the extent to which the data used to "train" the AI system may contain bias or discriminatory elements, and whether there are risks that the system that will operate on the basis of this data, as a result, will make biased decisions;

– analysis of risks at the stage of designing the AI system. Decisions made by the designers of the AI system may have significant consequences for human rights. Designers of such systems can, for example, prioritize the variables they would like the AI system to optimize and decide which variables the AI should take into account in its work. Such design decisions can have both positive and negative impacts on human rights, depending on the personal life experiences and biases of the designers;

– taking into account the algorithms of interactions: when an AI system is put into operation, it will interact with its environment in ways that may not have been intended at the initial stage. These interactions can have significant impacts on human rights. In some cases, the impact of these interactions can be detected using certain analytical techniques, but it is likely that some of the human rights implications of the use of AI systems will not be predictable using existing analytical mechanisms. This is not a unique problem that only concerns AI: pre-digital societies are extremely complex, and the impact of the actions of individuals and institutions on human rights cannot always be understood at the time of their production or for some time afterwards [8, p. 520].

Some scholars have highlighted the risks of unethical AI data replication. To the extent that AI accurately replicates past patterns of human decision-making, it inevitably continues to replicate social biases. Worse, unlike human decision-makers who have the free will to change their moral behavior over time, AI systems will not have such capabilities of their own in the foreseeable future. Instead, they will require constant intervention by those responsible for designing and operating such systems to ensure that their outcomes are consistent with evolving notions of fairness [9].

AI is inherently data-driven, largely based on algorithms that automatically analyze vast data sets to generate answers, predictions, and insights. Accordingly, AI systems rely on the collection, storage, consolidation, and analysis of large amounts of data. They also create strong incentives to collect and store as much additional data as possible, given the possibility that new data streams will allow AI systems to generate powerful new knowledge. Furthermore, even if techniques such as differential privacy are used to protect the privacy of individuals, AI technologies can generate information from such data that is then used to make predictions and act on the personal characteristics of a particular individual, while refraining from identifying the individual [10, p. 266].

In the context of the purpose of our study, we note that access to financial services, such as banking and credit, is an important means of promoting social and economic well-being. Access to credit, in particular, can now help disadvantaged individuals better enjoy their economic, social and cultural rights.

In recent years, lenders have begun to actively use AI to more accurately assess whether a potential borrower is a good credit risk. Unlike conventional credit scoring algorithms, the AI-based approach considers "all data as credit data" and analyzes a huge amount of data from many sources.

It should be noted that the use of artificial intelligence in financial institutions significantly expands the opportunities for the development of an ecosystem of online services, which in turn increase customer engagement and satisfaction. These services allow financial institutions to offer individualized solutions based on big data analysis, behavior prediction, and service automation. In Ukraine, where customers are increasingly focused on the quality and innovation of service, the use of AI in online services is of strategic importance. The ability of AI to analyze large volumes of transactional data allows us to identify customer behavior patterns. This contributes to the creation of personalized recommendations and special offers that best meet customer needs. For example, based on the analysis of payment or purchase history, the system can offer profitable financial products, such as loans or additional services that the customer will find useful [11].

The amount of data that AI-based credit scoring systems collect and analyze is alarming. One leading company in the US looks at over 3,000 variables when deciding whether to grant credit to an individual – including whether the individual is willing to take out credit for all areas and things at once, which is apparently associated with a higher risk of default. Another US company in the field examines an applicant's entire digital footprint, including social media usage, geolocation, website browsing habits, phone history (including text messages and call logs), shopping behavior, and more, to decide whether to extend their credit [12]. The use of AI in financial decision-making could even stifle freedom of thought, expression, and association, deterring people from engaging in activities they believe will negatively impact their credit score. This is not just a theoretical possibility. In 2009, American Express reduced the credit limit of an African-American businessman because other customers who had used his card at establishments where he had recently made purchases had a poor credit history [13].

Therefore, the introduction of AI into the lending process is likely to have a positive impact on the right to equality and non-discrimination for some individuals, while it will have a negative impact on others. On the positive side, the fact that AI-based algorithms take into account a wide range of

data sources could improve the ability of well-qualified individuals from marginalized communities to access credit. On the other hand, the specter of "network discrimination" that has a negative impact on the ability of members of these same communities to borrow money should not be underestimated. Finally, AI-based decision-making algorithms in the financial sector are likely to have a negative impact on freedom of belief and expression. In an era where "all data is credit data", people may feel restricted from expressing certain views or communicating with others for fear that an algorithm might assess their behavior as negative in a financial context.

One of the most significant factors in this process stems from the quality and accuracy of the data used to train these systems, as well as the fairness and accuracy of the data that these systems use to make decisions about a particular individual's loan application. Another stems from the subjective decisions that programmers make about how to code and classify the data they feed into their seemingly objective algorithms. In addition, the scores generated by AI can reproduce existing patterns of discrimination through "network discrimination." For example, if two individuals are in the same financial situation, the first person's friends live in "affluent" neighborhoods, while the other person's friends live in less "advantaged" neighborhoods. The algorithm may well determine that the first person is a better credit risk than the second. To the extent that such network factors correlate with classifications such as those based on race and gender, the potential for discriminatory effects is quite serious [14, p. 150].

5. Conclusions.

The rapid development and mass introduction of artificial intelligence technologies into various spheres of public life creates both unprecedented opportunities for improving the quality of human life and serious challenges for the system of human rights protection. Analysis of the impact of AI on human rights demonstrates the presence of a wide range of threats covering fundamental rights and freedoms: the right to non-discrimination (due to algorithmic bias), the right to privacy (due to the mass collection and processing of personal data), the right to freedom of expression (due to automated content moderation), the right to a fair trial (when using automated decision-making systems). Of particular danger are technologies of mass surveillance and biometric identification, autonomous weapons systems, as well as the use of deepfake technologies to manipulate public opinion, which undermines democratic processes and threatens the right to access reliable information.

The problem of ensuring the protection of human rights in the era of artificial intelligence is complicated by the specific characteristics of the functioning of AI systems, in particular their dependence on the quality of training data, the subjective decisions of developers at the design stage and the unpredictability of the consequences of algorithmic interactions. AI systems by their nature reproduce social biases embedded in the data of past human decision-making experience and, unlike people, do not have their own ability to change their behavior in accordance with the evolution of ethical norms of society. This creates risks of systematic discrimination of certain groups of the population and the consolidation of existing social inequalities through technological mechanisms that have the appearance of objectivity and neutrality.

The application of artificial intelligence in the financial sector, in particular in credit scoring systems, where algorithms analyze huge amounts of data about the applicant's digital footprint - from activity in social networks to geolocation and purchase history. While this approach may increase access to credit for some members of marginalized communities by taking into account alternative data sources, it also creates the risk of "network discrimination," where a person's financial capabilities are assessed based on the characteristics of their social environment. This not only violates the principle of individual responsibility, but can also limit freedom of belief and expression, as people feel pressure to self-censor, fearing the negative impact of their behavior on the algorithmic assessment of their creditworthiness.

These challenges highlight the need to develop a comprehensive regulatory framework for artificial intelligence at the national and international levels, which will ensure respect for human rights as a fundamental principle of the development and implementation of AI systems. Mechanisms to ensure transparency of algorithms, accountability of developers and users of AI systems, independent audits and assessments of human rights impacts, as well as effective procedures for challenging automated decisions and redress in cases of violations are critically important. International organizations,

including the Council of Europe, are already working to create legal standards in this area, but the success of these efforts will depend on the willingness of states to introduce strict requirements for compliance with human rights at all stages of the life cycle of artificial intelligence systems - from design to operation and decommissioning.

References:

1. Ombudsman of Ukraine. (2024). *Prava liudyny v epokhu shtuchnoho intelektu* [Human rights in the age of artificial intelligence]. https://ombudsman.gov.ua/storage/app/media/uploaded-files/%D0%9F%D0%A0%D0%90%D0%92%D0%90%20%D0%9B%D0%AE%D0%94%D0%98%D0%9D%D0%98%20%D0%92%20%D0%95%D0%9F%D0%9E%D0%A5%D0%A3%20%D0%A8%D0%A2%D0%A3%D0%A7%D0%9D%D0%9E%D0%93%D0%9E%20%D0%86%D0%9D%D0%A2%D0%95%D0%9B%D0%95%D0%9A%D0%A2%D0%A3_compressed.pdf [in Ukrainian]

2. Kravchuk, S.M. (2024). Vplyv shtuchnoho intelektu na prava liudyny ta zahalni rekomendatsii dlia staloho vtilennia [The impact of artificial intelligence on human rights and general recommendations for sustainable implementation]. *Visnyk Natsionalnoho universytetu «Lvivska politekhnika». Seriia: Yurydychni nauky* [Bulletin of Lviv Polytechnic National University. Series: Legal Sciences], 3(43), 101–110. [in Ukrainian]

3. Van Est, R., Gerritsen, J., & Kool, L. (2017). *Human rights in the robot age: Challenges arising from the use of robotics, artificial intelligence, and virtual and augmented reality* (Expert report written for the Committee on Culture, Science, Education and Media of the Parliamentary Assembly of the Council of Europe). Rathenau Instituut. https://www.rathenau.nl/en/publication/human-rights-robot-age-challengesarising-use-robotics-artificial-intelligence-and [in English]

4. Angwin, J., Larson, J., Mattu, S., & Kirchner, L. (2016). *Machine bias: There's software used across the country to predict future criminals. And it's biased against blacks*. ProPublica. https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing [in English]

5. Kravchuk, S.M. (2024). Vplyv shtuchnoho intelektu na prava liudyny ta zahalni rekomendatsii dlia staloho vtilennia [The impact of artificial intelligence on human rights and general recommendations for sustainable implementation]. *Visnyk Natsionalnoho universytetu «Lvivska politekhnika». Seriia: Yurydychni nauky* [Bulletin of Lviv Polytechnic National University. Series: Legal Sciences], 3(43), 105–112. [in Ukrainian]

6. Liu, H.-Y., & Zawieska, K. (2017). *A new human rights regime to address robotics and artificial intelligence*. http://www.werobot2017.com/wp-content/uploads/2017/03/Liu-A-NewHuman-Rights-Regime-to-Address-Robotics-and-Artificial-Intelligence.pdf [in English]

7. Raso, F. A., Hilligoss, H., Krishnamurthy, V., Bavitz, Ch., & Kim, L. (2018). *Artificial intelligence & human rights: Opportunities & risks* (Berkman Klein Center Research Publication No. 2018-6). https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3259344 [in English]

8. Russell, S. (2017). Artificial intelligence: The future is superintelligent. *Nature, 548*, 520–521. [in English]

9. Chander, A. (2017). The racist algorithm? *Michigan Law Review, 115*(6), 1023–1045. http://michiganlawreview.org/wp-content/uploads/2017/04/115MichLRev1023_Chander.pdf [in English]

10. Dwork, C., McSherry, F., Nissim, K., & Smith, A. (2006). Calibrating noise to sensitivity in private data analysis. In S. Halevi & T. Rabin (Eds.), *Theory of cryptography* (pp. 265–284). Springer Heidelberg. [in English]

11. Kozmenkov, M.H. (2025). Vykorystannia shtuchnoho intelektu v finansovykh ustanovakh v Ukraini: mozhlyvosti ta vyklyky dlia ekosystemy onlain servisiv [Use of artificial intelligence in financial institutions in Ukraine: Opportunities and challenges for the online services ecosystem]. *Problemy suchasnykh transformatsii. Seriia: ekonomika ta upravlinnia* [Problems of Modern Transformations. Series: Economics and Management], 19. https://file:///Users/a12345/Downloads/04_04_19_2025_Kozmenkov.pdf [in Ukrainian]

12. Lenddo. (2018, June 20). *Credit scoring: The LenddoScore fact sheet*. https://www.lenddo.com/pdfs/Lenddo_FS_CreditScoring_201705.pdf [in English]

13. Lieber, R. (2009, January 30). American Express watched where you shopped. *The New York Times*. https://www.nytimes.com/2009/01/31/your-money/credit-and-debit-cards/31money.html [in English]

14. Hurley, M., & Adebayo, J. (2016). Credit scoring in the era of big data. *Yale Journal of Law & Technology, 18*(1), 148–216.

15. Bielova, M.V., & Bielov, 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]

Terezia Popovich,
Candidate of Law, Associate Professor,
Associate Professor of the Department of Theory and History of State and Law
State Higher Educational Institution "Uzhgorod National University"
ORCID: 0000-0002-8333-3921