

D. S. DRAKON

Higher Education Student
ERI “Karazin Banking Institute”
of V. N. Karazin Kharkiv National University
ORCID: 0009-0004-3712-7483

N. I. STIAHLYK

Ph.D., Head of the Department of Information Technology
and Mathematical Modeling
ERI “Karazin Banking Institute”
of V. N. Karazin Kharkiv National University
ORCID: 0000-0002-0573-3508

THE ETHICS OF ARTIFICIAL INTELLIGENCE: BOUNDARIES AND RESPONSIBILITY

Artificial intelligence is increasingly integrating into everyday life; however, its use is accompanied by a significant number of ethical challenges. The lack of algorithmic transparency, potential discrimination, risks associated with military applications, and issues of responsibility for autonomous decisions require thorough analysis. This study examines approaches to AI regulation worldwide, particularly in the European Union, the United States, China, and Ukraine. It has been established that the EU applies strict regulatory frameworks to protect human rights, the US favours self-regulation by AI development companies, China enforces state control over AI development and implementation, which raises concerns regarding the protection of citizens' rights, while Ukraine is only beginning to establish its own regulatory framework.

The research identifies the fundamental ethical principles that should guide AI usage and, based on these, proposes recommendations for regulating the technology. These include the necessity of algorithmic transparency, personal data protection, the right to be forgotten, restrictions on autonomous decision-making in critical areas, and the establishment of international standards. The study analyses the prospects for regulatory development and highlights the importance of balancing technological progress with oversight to prevent the uncontrolled use of AI.

Additionally, issues of responsibility for decisions made by artificial intelligence and the necessity of independent oversight over its implementation were considered. The study examines possible scenarios for technological development under global standardisation, as well as the risks that may arise from insufficient regulation. The findings demonstrate that only a comprehensive approach to regulation and accountability will help mitigate potential threats and ensure the safe development of AI for society, minimising the risks associated with its uncontrolled application.

Key words: artificial intelligence, AI ethics, algorithmic transparency, autonomous systems, digital rights, AI regulation, responsibility, innovative technologies, technology oversight.

Д. С. ДРАКОН

здобувач вищої освіти
Навчально-науковий інститут «Каразінський банківський інститут»
Харківського національного університету імені В. Н. Каразіна
ORCID: 0009-0004-3712-7483

Н. І. СТЯГЛИК

кандидат технічних наук,
завідувач кафедри інформаційних технологій
та математичного моделювання
Навчально-науковий інститут «Каразінський банківський інститут»
Харківського національного університету імені В. Н. Каразіна
ORCID: 0000-0002-0573-3508

ЕТИКА ШТУЧНОГО ІНТЕЛЕКТУ: МЕЖІ ТА ВІДПОВІДАЛЬНІСТЬ

Штучний інтелект (ШІ) стрімко розвивається та проникає в різні сфери суспільного життя, відкриваючи нові можливості, але водночас породжуючи серйозні етичні виклики. Його застосування у медицині, правосудді, безпеці, бізнесі та інших галузях вимагає не лише технологічного вдосконалення, а й ефективного правового регулювання для запобігання дискримінації, порушення прав людини та зловживання технологіями. У статті розглянуто основні етичні та моральні проблеми, які пов'язані із використанням інструментів штучного

інтелекту. Проведено аналіз підходів до регулювання ШІ в Європейському Союзі, США, Китаї та Україні, які демонструють різний рівень державного контролю та саморегулювання.

Визначено ключові етичні принципи використання ШІ, зокрема прозорість алгоритмів, захист персональних даних, відповідальність за автономні рішення та необхідність міжнародних стандартів. Окрема увага приділена питанням контролю за розвитком ШІ, включно з проблемами упередженості алгоритмів, безпеки персональних даних і потенційних загроз, пов'язаних із використанням ШІ в автономних системах та військовій сфері. Також досліджено перспективи глобальної стандартизації, виклики, що постають перед державами та корпораціями, а також ризики, пов'язані з недостатнім регулюванням технологій.

На основі аналізу світового досвіду сформульовано рекомендації щодо регулювання ШІ, спрямовані на забезпечення його етичного розвитку та балансування між технологічним прогресом і безпекою суспільства. Підкреслено необхідність міжнародної співпраці у створенні єдиних стандартів, які допоможуть мінімізувати ризики неконтрольованого розвитку ШІ та гарантувати його відповідальне впровадження в усі сфери життя.

Ключові слова: штучний інтелект, етика ШІ, прозорість алгоритмів, автономні системи, цифрові права, регулювання ШІ, відповідальність, інноваційні технології, контроль технологій.

Problem Statement

Artificial Intelligence (hereinafter AI) is a technology that enables machines and software systems to perform tasks traditionally considered the domain of human intelligence: pattern recognition, decision-making, natural language processing, learning from large volumes of data, and emotional responsiveness. At present, AI is widely used in many areas of our society, including: business, which encompasses data analytics, automation, and financial monitoring; healthcare, which includes disease diagnosis and personalised medicine; law, which involves the automation of legal documents, legislative analysis, and the completion of legal forms; transport, particularly autonomous vehicles, route optimisation, and advanced traffic systems; security, which covers facial and fingerprint recognition systems, crime prediction, and an advanced level of security system protection, among others. AI has also extended into personal use by individual users for time and financial planning, receiving advice, conducting analysis and task execution, generating content for various fields, and even engaging in ordinary conversations.

However, the rapid development and broad scope of AI applications are accompanied by ethical challenges that are becoming increasingly evident and require discussion, particularly in defining the limits of its accessibility. This is due to the fact that AI is becoming increasingly autonomous and is beginning to influence the process of making critically important decisions.

Analysis of Recent Research and Publications

Most scholars focus on the risks and benefits of autonomous systems, moral aspects, issues of discrimination, and ensuring transparency in AI decision-making. For instance, S. Russell, in Human Compatibility [1, p. 7], explores the concept of developing technologies that prioritise human interests. Meanwhile, Jean-François Bonnefon, as part of the Moral Machine project [2], analyses how people perceive ethical considerations in autonomous vehicles, examining the criteria such systems should follow.

Despite the numerous ethical dilemmas, contemporary research in AI ethics covers a broad range of issues and challenges associated with the implementation of this technology.

Research Aim

The aim of this study is to analyse the ethical aspects of AI usage, particularly its impact on society, the responsibility of developers, and regulatory frameworks. Special attention will be given to assessing the extent to which ethical principles are considered in the development and implementation of this technology. The research is focused on identifying the key risks associated with AI usage and determining ways to mitigate them through legal, technical, and organizational measures.

The main research objectives are:

- Analysis of the current state of ethical standards for AI and their implementation in international practice;
- Identification of key issues related to AI autonomy, algorithmic transparency, and accountability for AI-driven decisions;
- Examination of global AI regulatory frameworks, with a focus on the most effective practices;
- Development of recommendations for the integration of ethical norms into AI development and application to ensure its safe and fair use;
- A Analysis of international AI regulation initiatives and assessment of their effectiveness in addressing ethical challenges, particularly in preventing discrimination, bias, and algorithmic opacity;
- Identification of key directions for the development of ethical standards for AI.

Presentation of the Main Material

Let us outline the key reasons for the increasing ethical concerns surrounding the use of AI tools. First and foremost, automated decisions can be unfair, as system algorithms may contain hidden biases and discriminate against individuals based on various factors or, conversely, fail to consider certain significant aspects of information. Additionally, AI operates

with a lack of transparency, evidenced by the fact that many neural network decisions are difficult to explain even to their developers (the so-called “black box” problem). Moreover, AI’s processing of vast amounts of personal data creates a risk of leaks and misuse, posing threats to data privacy and security.

Another critical issue is the lack of accountability due to the use of autonomous systems – there is no specific individual responsible for an error made by an autonomous vehicle or a medical algorithm. Furthermore, the use of AI in military applications raises serious moral concerns, particularly in the development of autonomous weaponry without human oversight.

Several fundamental questions arise: Does AI even comprehend ethical dilemmas and moral aspects? Can AI make morally justified decisions in the same way a human does?

Since AI operates based on algorithms and statistical models, lacking consciousness, empathy, and moral awareness, it can analyse vast amounts of data, imitate ethical decision-making, and even exhibit behaviour that resembles human empathy. However, AI does not possess its own moral compass – it merely follows predefined criteria, does not fully grasp the consequences of its decisions, and simply selects the most optimal option from those available. Its decisions are entirely dependent on data – if trained on biased datasets, it will replicate the same mistakes.

The development of AI raises numerous ethical and moral concerns, necessitating clear ethical and legal frameworks. If society fails to establish boundaries for the application of these technologies, it may lead to severe consequences, ranging from human rights violations to the emergence of uncontrolled autonomous systems. Therefore, AI ethics remains a key challenge for modern society.

Among the most pressing issues that require resolution are: how to legally regulate AI, who should be held accountable for its mistakes and biases, and which areas of AI application require the strictest oversight.

Finding answers to these questions and establishing ethical standards and mechanisms of accountability should serve as the foundation for the further development of AI, aiming to minimise risks and ensure its safe future use.

In addition to scientific research, major technology companies play a significant role in shaping AI ethical principles. Companies such as Google, OpenAI, and IBM have developed their own frameworks for the responsible use of artificial intelligence, focusing on aspects such as transparency, fairness, and security.

Google has formulated a set of AI principles, known as AI Principles [3], which declare the responsibility of developers towards society. IBM, in turn, has introduced the concept of Explainable AI [4], which aims to enhance the understanding of how neural networks make decisions. OpenAI has launched a research initiative dedicated to the safe development of artificial general intelligence (AGI), emphasising the importance of global cooperation in minimising risks [10]. Microsoft is also actively involved in shaping ethical standards by implementing its Responsible AI policy and working on the development of fairness-oriented, non-discriminatory algorithms.

However, some experts criticise these companies for their declarative commitments to ethical AI, arguing that they are often not accompanied by tangible changes in the development approaches of their technologies.

Furthermore, to regulate AI usage, several international organisations have formulated their principles for the ethical application of AI technologies. UNESCO, in its Recommendation on the Ethics of AI [5], highlights the necessity of ensuring equal access to technology and protecting human rights, while the European Union’s Ethics Guidelines for Trustworthy AI [6] focus on transparency, fairness, and algorithmic accountability. IEEE has also developed the Ethically Aligned Design document [7], which provides recommendations for AI developers, particularly in avoiding discrimination and ensuring the explainability of AI-driven decisions.

Despite the existence of these initiatives, most of these principles remain advisory, and their implementation largely depends on the policies of individual countries and companies. The absence of a unified global AI regulation mechanism threatens the consistent enforcement of ethical standards across all AI applications.

For this reason, the following sections of this study will focus on analysing legislative initiatives in various countries and efforts to standardise the use of artificial intelligence.

Different countries have distinct approaches to AI regulation, reflecting their political, economic, and social characteristics. Let us examine these approaches using the examples of the EU, the US, China, and Ukraine.

The European Union has focused on protecting human rights, ensuring the ethical use of technology, and implementing legal oversight. Its goal is to guarantee the safe development of AI and prevent misuse by users. In April 2019, the European Commission published the Ethics Guidelines for Trustworthy AI [6], which outlined seven key requirements: human oversight, technical robustness, privacy and data governance, transparency, diversity, non-discrimination and fairness, societal well-being, and accountability. However, the primary regulatory document developed by the EU to govern AI is the Artificial Intelligence Act (AI Act) [11], which was adopted by the European Parliament in June 2024.

According to the key provisions of the AI Act, AI systems are classified based on their risk levels into four categories: minimal risk (such as chatbots and recommendation systems), limited risk (AI in financial services and recruitment), high risk (AI in critical infrastructure, legal systems, security systems), and unacceptable risk (AI for mass social scoring of citizens, subliminal manipulation, political propaganda). Transparency and accountability provisions require companies developing high-risk AI systems to provide detailed explanations of their operation, test them for compliance with ethical

standards, and implement safety assessment mechanisms. Failure to comply with these requirements can result in fines of up to €30 million or 6 % of the company's global annual revenue.

The EU opts for strict regulation to minimise risks; however, this approach may slow down innovation due to complex bureaucratic procedures and high compliance requirements for AI developers. According to the General Data Protection Regulation (GDPR), the minimum age for providing consent for personal data processing in most EU countries is 16 years. This means that individuals under 16 require parental or guardian consent to use such services. As a result, many companies specify in their policies that users must be at least 16 years old if they are citizens or residents of the EU.

In contrast, the United States does not have a single, nationwide AI regulation law. Instead, states adopt a flexible self-regulation model, where the primary responsibility for adhering to ethical standards lies with private companies and industry-specific standards. In October 2023, the White House published the Blueprint for an AI Bill of Rights [12], which outlines five principles: safe and effective systems, protection against algorithmic discrimination, data privacy, notice and explanation, and human alternatives, consideration, and fallback options. Moreover, due to minimal government intervention in the tech sector in the US, AI standards are primarily shaped by leading tech companies such as Google, Microsoft, and OpenAI. However, while the American model encourages rapid AI development with minimal government interference, it does not guarantee sufficient user protection and raises concerns about the potential for AI-driven public opinion manipulation and mass surveillance.

The Chinese system is the strictest in the world, as the government actively controls information technologies, employs AI for state surveillance, and creates isolated networks. Some of China's key AI regulations include licensing requirements for AI-developing companies (all algorithms must be approved by the government), strict content censorship, social credit systems using AI to assess "public behaviour," and restrictions on gaming and video content. Additionally, China is actively developing military AI technologies, including autonomous drones, intelligence systems, and cyberattack capabilities.

Ukraine does not have a dedicated law regulating AI; however, it is actively working on developing legislative initiatives in this field. Despite the absence of a specific AI law, certain legal and regulatory frameworks already exist in Ukraine to govern AI usage. These include documents such as the Law of Ukraine on Personal Data Protection, the Digital Transformation Strategy of Ukraine, and the National AI Development Strategy.

In 2020, the Ministry of Digital Transformation of Ukraine introduced the Concept for AI Development in Ukraine until 2030, which outlines the goals, principles, and objectives for AI technology advancement in the country. The document emphasises human capital development, support for entrepreneurship, AI implementation in the public sector, and ensuring ethical and legal aspects of AI usage. To implement this concept, in 2021, an action plan was approved, which includes specific steps and projects for integrating AI into various sectors of society.

As a country striving for European Union integration, Ukraine must work even more actively on AI legislation and align its policies with the strict regulatory framework of the EU. However, this process is complicated by several factors, including martial law, insufficient funding, and brain drain. Despite these challenges, Ukraine has taken significant steps towards AI regulation recently by initiating several studies and proposals. These include the Recommendations of the National Security and Defense Council of Ukraine on AI Usage in Military and Cybersecurity, cooperation with the EU within the "Digital Ukraine" initiative, and parliamentary projects aimed at establishing AI-related regulations for developers and users.

This indicates that Ukraine is currently in the active phase of shaping its approach to AI regulation.

By analysing the experience of leading technological countries in AI regulation and considering the specifics of AI implementation in Ukraine, key recommendations can be formulated for the ethical use of AI:

- *Establishing restrictions on autonomous decision-making*, which involves defining areas where AI cannot make decisions without human oversight and implementing a "human control" mechanism to ensure accountability.
- *Ensuring algorithmic transparency and accountability*, requiring developers to provide clear explanations of how AI systems operate and implementing independent audits to identify potential discrimination and errors.
- *Protecting personal data*, which includes strict regulations on the use of personal data, the establishment of a «right to be forgotten» principle allowing users to delete their data from AI algorithms, and ensuring the secure storage of such data.
- *Limiting AI usage in military and governmental sectors*, which entails prohibiting AI from being used in uncontrolled autonomous weapons and establishing independent ethical committees to assess AI applications in these fields.

Furthermore, There Are Potential Prospects for the Development of Ethical AI Regulation:

- Development of unified AI regulation standards through cooperation between international organisations.
- Creating a balance between innovation and AI oversight.
- Enhancing digital literacy in AI.
- Implementing the principle of "ethics by default".
- Preventing negative AI development scenarios.

Conclusions

AI is one of the most influential technologies of our time, with capabilities that are both impressive and alarming. It is actively transforming various sectors of society – from business and medicine to law and public administration. However, its rapid development is accompanied by a range of ethical challenges that require immediate resolution. The analysis

of international experience has demonstrated different approaches to AI regulation, ranging from strict state control to flexible self-regulation models.

It has been determined that for the further ethical development of AI, the following measures must be implemented: restrictions on autonomous decision-making in critical areas, transparency, and explainability of algorithms, protection of personal data and the “right to be forgotten,” establishment of ethical committees, a balance between progress and control, and ensuring the responsible use of AI.

“If society does not establish clear ethical boundaries and fails to develop an effective system of AI oversight, this could lead to severe consequences-ranging from discrimination and human rights violations to uncontrolled autonomous systems. Thus, the rise of the machines is not an inevitability, but merely a scenario that can be avoided through proactive and responsible action.” – (ChatGPT, 2025).

Bibliography

1. Рассел, С. Human Compatible: Artificial Intelligence and the Problem of Control / Стюарт Рассел. Нью-Йорк : Viking, 2019. 352 с.
2. Bonnefon, J.-F., Shariff, A., Rahwan, I. Moral Machine [Електронний ресурс] / Massachusetts Institute of Technology, 2018. Режим доступу: <https://www.moralmachine.net/>
3. Google. AI Principles [Електронний ресурс] / Google, 2018. Режим доступу: <https://ai.google/principles/>
4. IBM. Explainable AI [Електронний ресурс] / IBM, 2021. Режим доступу: <https://www.ibm.com/watson/explainable-ai>
5. UNESCO. Recommendation on the Ethics of Artificial Intelligence [Електронний ресурс] / UNESCO, 2021. Париж. Режим доступу: <https://unesdoc.unesco.org/ark:/48223/pf0000380455>
6. European Commission. Ethics Guidelines for Trustworthy AI [Електронний ресурс] / European Commission, 2019. Режим доступу: <https://ec.europa.eu/digital-strategy/en/news/ethics-guidelines-trustworthy-ai>
7. IEEE. Ethically Aligned Design [Електронний ресурс] / IEEE, 2019. Режим доступу: <https://ethicsinaction.ieee.org/>
8. OpenAI. OpenAI Charter [Електронний ресурс] / OpenAI, 2018. Режим доступу: <https://openai.com/charter/>
9. OpenAI. Artificial General Intelligence and the Future of Humanity [Електронний ресурс] / OpenAI, 2021. Режим доступу: <https://openai.com/research/>
10. Microsoft. Responsible AI Principles and Approach [Електронний ресурс] / Microsoft Corporation, 2022. Режим доступу: <https://www.microsoft.com/en-us/ai/principles-and-approach>
11. European Commission. Artificial Intelligence Act [Електронний ресурс] / European Commission, 2021. Режим доступу: <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A52021PC0206>
12. The White House. Blueprint for an AI Bill of Rights [Електронний ресурс] / The White House, Office of Science and Technology Policy (OSTP), 2022. Режим доступу: <https://bidenwhitehouse.archives.gov/ostp/ai-bill-of-rights/>
13. ChatGPT. Ethics of Artificial Intelligence: Limits and Responsibility [Електронний ресурс] / OpenAI, 2024. Режим доступу: <https://openai.com>

References

1. Russell, S. (2019). *Human compatible: Artificial intelligence and the problem of control*. Viking.
2. Bonnefon, J.-F., Shariff, A., & Rahwan, I. (2018). *Moral Machine*. Massachusetts Institute of Technology. Retrieved February 21, 2025, from <https://www.moralmachine.net/>
3. Google. (2018). *AI principles*. Retrieved February 21, 2025, from <https://ai.google/principles/>
4. IBM. (2021). *Explainable AI*. Retrieved February 21, 2025, from <https://www.ibm.com/watson/explainable-ai>
5. UNESCO. (2021). *Recommendation on the ethics of artificial intelligence*. Paris. Retrieved February 20, 2025, from <https://unesdoc.unesco.org/ark:/48223/pf0000380455>
6. European Commission. (2019). *Ethics guidelines for trustworthy AI*. Retrieved February 16, 2025, from <https://ec.europa.eu/digital-strategy/en/news/ethics-guidelines-trustworthy-ai>
7. IEEE. (2019). *Ethically aligned design*. Retrieved February 17, 2025, from <https://ethicsinaction.ieee.org/>
8. OpenAI. (2018). *OpenAI charter*. Retrieved February 22, 2025, from <https://openai.com/charter/>
9. OpenAI. (2021). *Artificial general intelligence and the future of humanity*. Retrieved February 15, 2025, from <https://openai.com/research/>
10. Microsoft. (2022). *Responsible AI principles and approach*. Retrieved February 17, 2025, from <https://www.microsoft.com/en-us/ai/principles-and-approach>
11. European Commission. (2021). *Artificial Intelligence Act*. Retrieved February 17, 2025, from <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A52021PC0206>
12. The White House, Office of Science and Technology Policy (OSTP). (2022). *Blueprint for an AI bill of rights*. Retrieved February 20, 2025, from <https://bidenwhitehouse.archives.gov/ostp/ai-bill-of-rights/>
13. ChatGPT. (2024). *Ethics of artificial intelligence: Limits and responsibility*. OpenAI. Retrieved February 17, 2025, from <https://openai.com>