

UDK: 004.8 : 34 (4-6 EU)

Review paper

Krzysztof Bodzenta, Teaching Assistant

Krakow University of Economics, Poland

E-mail: bodzentk@uek.krakow.pl

ORCID ID: 0009-0004-2151-6134

THE ANTHROPOCENTRIC PARADIGM OF ARTIFICIAL INTELLIGENCE REGULATION IN EUROPEAN UNION LAW*

ABSTRACT

This article examines the specific characteristics of the European Union's regulatory framework for artificial intelligence (AI), which stands out among global strategies due to its human-centric approach. It first focuses on conceptualizing the notion of "artificial intelligence". Subsequently, it identifies the sources of risk and potential threats associated with AI applications that justify the need for its regulation. The next section analyzes the fundamental principles of the EU regulatory model, which is based on a balanced approach and the promotion of trustworthy AI. Furthermore, the article explores the key legal mechanism introduced under the AI Act to safeguard values and fundamental rights deemed particularly significant by the EU legislator. Finally, the article presents concluding remarks on the European Union's approach to AI regulation.

Keywords: *Artificial Intelligence, AI Act, European Union, Public Regulation of AI, European approach to AI*

*In short, success in creating AI, could be the biggest event in the history of our
civilisation.*

But it could also be the last, unless we learn how to avoid the risks.

Stephen Hawking¹

* The article presents the result of the Project no 095/EPG/2024/POT financed from the subsidy granted to the Krakow University of Economics.

¹ This quote comes from a speech delivered at the inauguration of the Centre for the Future of Intelligence at the University of Cambridge on 19 October 2016. See University of Cambridge, The Best or Worst Thing to Happen to Humanity? Stephen Hawking Launches Centre for the Future of Intelligence, <https://www.cam.ac.uk/research/news/the-best-or-worst-thing-to-happen-to-humanity-stephen-hawking-launches-centre-for-the-future-of>, 13 April, 2025.

1. INTRODUCTION

Artificial Intelligence (hereinafter AI) is no longer a technology of the future - it has become a technology of the present. Modern AI systems operate in the background of nearly every aspect of our lives, often influencing decisions that affect us without our direct awareness. They recommend movies and music on streaming platforms, manage urban traffic, support medical diagnostics, analyze financial data, and assist in automating business processes. AI is integrated into our smartphones, cars, banking systems, e-commerce, and even public administration.

Thanks to the rapid advancement of machine learning algorithms and increasing computational power, AI has become not only ubiquitous but also increasingly sophisticated. Generative models, such as chatbots and content creation systems, are transforming communication and access to information, while autonomous systems can make decisions in critical sectors such as healthcare, finance, and the justice system.

However, the expanded use of AI brings not only significant benefits but also new challenges and risks. Concerns arise regarding privacy, algorithmic transparency, the impact of AI on the job market, and the ethical implications of its application. Decisions made by algorithms can be biased, difficult to interpret, and even discriminatory. Moreover, AI can become a tool for information manipulation, surveillance, and social control. The growing significance of AI and the risks associated with it have naturally attracted the attention of legislators worldwide. In response to the challenges posed by AI, individual countries and international organizations have adopted different approaches to its regulation.

On a global scale, three major players in the field of AI stand out - the United States (hereinafter the US), China, and the European Union (hereinafter the EU) - each of which has developed its own regulatory model for this technology. The model adopted in the US is limited to sector-specific regulations, relying on a free-market approach, as the country prioritizes stimulating the development of innovative technologies and maintaining a competitive edge. In the case of China, AI regulations are closely tied to state-driven policy discourse, with the primary objective of ensuring control over the development and use of AI.

The EU takes a fundamentally different approach, being the first in the world to establish comprehensive, horizontal legal frameworks for AI - Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonised rules on artificial intelligence and amending Regulations (EC) No 300/2008, (EU) No 167/2013, (EU) No 168/2013, (EU) 2018/858, (EU)

2018/1139, and (EU) 2019/2144, as well as Directives 2014/90/EU, (EU) 2016/797, and (EU) 2020/1828, OJ L 2024/1689 of 12/7/2024 (hereinafter referred to as the AI Act or Regulation). The adopted regulations focus on promoting trustworthy, human-centric AI, balancing the benefits of technological advancement with the need to protect individual rights and safeguard the interests of society as a whole.

This article focuses on the third approach - the EU model of AI regulation, which establishes an anthropocentric paradigm in the public law regulation of this technology. The aim of the study is to provide an overview of this regulatory framework, with particular emphasis on protected rights and the legal mechanism ensuring their enforcement.

2. WHAT IS AI? IN SEARCH OF A UNIFIED DEFINITION

From the perspective of the historical development of cybernetic technologies, AI is not a new invention. Its origins date back to the second half of the 20th century when, alongside advancements in computer science, the first research began on algorithms enabling machines to mimic human thought processes. The creation of the term "artificial intelligence" is attributed to John McCarthy,² who defined it as an interdisciplinary field of science and engineering focused on constructing systems capable of exhibiting intelligent³ behaviors characteristic of humans.⁴ For decades, AI was primarily developed in "laboratory" conditions, remaining a subject of interest mainly for specialists in computer science, mathematics, engineering, and cognitive science. A breakthrough occurred at the turn of the 20th and 21st centuries when the rapid advancement of neural networks and machine learning algorithms led to the emergence of systems capable of independently processing information and making decisions. AI began to move beyond the realm of research and experimentation, entering sectors such as the economy, public administration, and everyday life. The current abundance of AI-based solutions

² The term "artificial intelligence" is believed to have been first used by John McCarthy during the Dartmouth conference in 1956. See McCarthy, J.; Minsky, M.L.; Rochester, N.; Shannon, C.E., *A Proposal for the Dartmouth Summer Research Project on Artificial Intelligence*, August 1955, <http://jmc.stanford.edu/articles/dartmouth/dartmouth.pdf>, 13 April, 2025.

³ According to the Oxford English Dictionary, intelligence is defined as "the faculty of understanding". See Oxford English Dictionary, https://www.oed.com/dictionary/intelligence_n, 13 April, 2025.

⁴ See McCarthy, J., What is Artificial Intelligence?, November 2007, p. 2, <https://www-formal.stanford.edu/jmc/whatisai.pdf>, 13 April, 2025. This understanding of AI emerged in the early 1950s and is associated with the Turing test, which aimed to assess whether a machine could exhibit intelligent behavior indistinguishable from that of a human.

and its dynamic development make the task of precisely defining this concept extremely challenging, if not impossible. The approach to defining AI varies depending on the context- its meaning differs in everyday language, in scientific and technological environments, and within legal discourse.

From the perspective of this analysis, the definition proposed by the EU, developed for the regulation of this technology, is of particular importance. The process of shaping the EU's definition of AI was gradual. One of the first definitions included in official EU documents was proposed by the European Commission in the 2018 communication *Artificial Intelligence for Europe*.⁵ It stated that: "Artificial intelligence (AI) refers to systems that display intelligent behaviour by analysing their environment and taking actions – with some degree of autonomy – to achieve specific goals. AI-based systems can be purely software-based, acting in the virtual world (e.g. voice assistants, image analysis software, search engines, speech and face recognition systems) or AI can be embedded in hardware devices (e.g. advanced robots, autonomous cars, drones or Internet of Things applications).

Although this definition served as a starting point for further legislative work, it remained too general. It did not specify concepts such as "intelligent behaviour", and the examples of AI applications it provided did not fully reflect the broad spectrum of this technology's capabilities.⁶

⁵ The European Commission, Communication from the Commission to the European Parliament, the European Council, the Council, the European Economic and Social Committee and the Committee for the Regions, *Artificial Intelligence for Europe*, COM(2018) 237 final, Brussels, 25 April 2018, <https://eur-lex.europa.eu/legal-content/PL/TXT/PDF/?uri=CELEX:52018DC0237>, 13 April, 2025.

⁶ It is worth mentioning that the High-Level Expert Group on Artificial Intelligence (HLEG), established by the European Commission, also contributed to the development of an AI definition. The group proposed the following definition: "Artificial intelligence (AI) systems are software (and possibly also hardware) systems designed by humans that, given a complex goal, act in the physical or digital dimension by perceiving their environment through data acquisition, interpreting the collected structured or unstructured data, reasoning on the knowledge, or processing the information, derived from this data and deciding the best action(s) to take to achieve the given goal. AI systems can either use symbolic rules or learn a numeric model, and they can also adapt their behaviour by analysing how the environment is affected by their previous actions. As a scientific discipline, AI includes several approaches and techniques, such as machine learning (of which deep learning and reinforcement learning are specific examples), machine reasoning (which includes planning, scheduling, knowledge representation and reasoning, search, and optimization), and robotics (which includes control, perception, sensors and actuators, as well as the integration of all other techniques into cyber-physical systems)". See the High-Level Expert Group (European Commission) on Artificial Intelligence, *Ethics Guidelines for Trustworthy AI*, April 2019, p. 8, https://ec.europa.eu/newsroom/dae/document.cfm?doc_id=60651, 13 April, 2025.

It is worth noting that the further development of the EU's conceptual framework for AI was also significantly influenced by work carried out at the level of the Organisation for Economic Co-operation and Development (OECD). In its 2019 recommendation, the OECD introduced a more precise definition of AI systems as: "machine-based systems that, for explicit or implicit objectives, infer, from the input they receive, how to generate outputs such as predictions, content, recommendations, or decisions that can influence physical or virtual environments. Different AI systems vary in their levels of autonomy and adaptiveness after deployment."⁷

This definition became a reference point for the further work of the EU legislator, who adopted the following legal definition in the final version of the Regulation: "AI system means a machine-based system that is designed to operate with varying levels of autonomy and that may exhibit adaptiveness after deployment, and that, for explicit or implicit objectives, infers, from the input it receives, how to generate outputs such as predictions, content, recommendations, or decisions that can influence physical or virtual environments."⁸

In the current landscape and in light of projected AI development trends, this definition appears to be the most comprehensive, while also remaining flexible and resilient to future technological changes. It highlights the key characteristics of AI systems, particularly their ability to operate autonomously and adapt after deployment, distinguishing them from traditional, static systems. Another crucial aspect is the recognition of the impact AI-driven decisions can have on both virtual and real-world environments. This broad conceptualization allows for an extensive interpretation of AI applications - ranging from predictive and recommendation systems to autonomous machines and analytical tools - whose functioning may have social, economic, or legal implications.

For the purposes of this article, AI will be understood in accordance with the definition adopted in the EU Regulation.

3. THE NEED FOR REGULATING AI

Undoubtedly, the need for AI regulation has arisen due to its rapid development and the simultaneous increase in risks associated with its expanding application, which - if left unchecked - could lead to serious negative consequences in the long

⁷ OECD Recommendation of the Council on Artificial Intelligence, 22 May 2019, <https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0449>, 13 April, 2025.

⁸ Article 3 item (1) of the AI Act.

run. The spectrum of potential risk sources and resulting threats related to AI usage is vast, stemming from various domains.⁹ Therefore, it is challenging to comprehensively address this issue within the scope of a single article. Nevertheless, it is worth examining the key aspects of this topic. To structure this discussion, four main areas of risk can be distinguished:

1. Technical Risks – these stem primarily from the complexity and technological limitations of AI systems. Examples include:
 - unpredictability and the “black box” problem – AI models, especially those based on machine learning, may produce unexpected results, with decision-making mechanisms difficult to interpret even for their creators;¹⁰
 - algorithmic errors and model drift – AI models can lose accuracy over time, particularly when the data they rely on changes, leading to increasingly flawed outputs;¹¹

⁹ To illustrate this scale, it is worth referring to the findings presented in the study by Slattery, P.; Saeri, A. K.; Grundy, E. A. C.; Graham, J.; Noetel, M.; Uuk, R.; Dao, J.; Pour, S.; Casper, S.; Thompson, N., *The AI Risk Repository: A Comprehensive Meta-Review, Database, and Taxonomy of Risks From Artificial Intelligence*, August 2024, <http://dx.doi.org/10.13140/RG.2.2.28850.00968>, 13 April, 2025. This study constitutes a meta-analysis of 43 existing classifications of AI-related risks and identifies 777 unique threats. The key risk categories include discrimination and toxicity, privacy and security violations, disinformation, malicious activities, challenges in human-computer interaction, socio-economic and environmental risks, as well as failures and technological limitations of AI systems.

¹⁰ Although the architecture of machine learning-based systems (such as neural networks or decision trees) is generally transparent, the learning process itself means that programmers have no control over the specific parameter values, which are autonomously developed by the algorithms based on training data. As a result, these models learn complex conceptual representations—often at a level of abstraction that has no direct equivalent in the model’s structure. Additionally, internal parameters can interact in nonlinear and multilayered ways, making it extremely difficult to fully interpret the relationships between input data and the model’s outputs. Even the creators of such systems may be unable to precisely determine how a given result was produced or explain why the model made a particular decision. For this reason, such systems are often referred to as “black boxes” — their internal logic remains opaque, despite their operational effectiveness. See Zednik, C., *Solving the Black Box Problem: A Normative Framework for Explainable Artificial Intelligence*, March 2019, Philosophy & Technology, p. 4-6, <https://doi.org/10.48550/arXiv.1903.04361>, 13 April, 2025.

¹¹ One practical manifestation of algorithmic errors in artificial intelligence models is the phenomenon of so-called “hallucinations” – instances where an AI system generates content that appears coherent and plausible, but is in fact entirely false or nonexistent, such as fabricated data, quotes, sources, or references. This problem is particularly prevalent in large language models, which rely on statistical relationships between words without genuinely “understanding” the truthfulness of the information they produce. For example, a study conducted in 2023 aimed to assess the reliability of bibliographic references generated by ChatGPT in academic texts. The research team examined 178 generated citations, of which as many as 39% contained incorrect or non-existent DOI identifiers, while an additional 16% referred to publications that could not be located using available

- sensitivity to input data – AI performance may be compromised if it receives incorrect or manipulated data (e.g., adversarial attacks¹²).
2. Social Risks – these arise from AI's impact on society, such as:
 - lack of transparency and technological complexity – The advanced nature of AI systems makes them difficult for the average user to understand, lowering public trust in AI-driven decisions;
 - exploitation of AI for disinformation and manipulation – AI algorithms can be misused to deceive the public, generate fake content (e.g., deepfakes), or personalize information in a way that fosters social polarization through “filter bubbles”;¹³
 - workforce automation and labor market shifts – AI-driven automation in key industries may lead to job losses, increasing social tensions and economic inequalities.¹⁴
 3. Legal Risks – these result from inadequate regulatory frameworks and delays in AI-specific legislation:

databases. See Athaluri, S.; Manthena, S.; Kesapragada, V.; Yarlagadda V.; Dave, T.; Duddumpudi, R., *Exploring the Boundaries of Reality: Investigating the Phenomenon of Artificial Intelligence Hallucination in Scientific Writing Through ChatGPT References*, April 2023, Cureus, <https://doi.org/10.7759/cureus.37432>, 13 April, 2025.

¹² Adversarial attacks are intentional manipulations of input data designed to cause AI systems to make incorrect or misleading decisions. The literature distinguishes several main types of such attacks, including: evasion attacks (modifying input data at inference time to mislead the system), poisoning attacks (corrupting training datasets to insert vulnerabilities), backdoor attacks (embedding hidden triggers that activate malicious behavior), model extraction attacks (attempts to reconstruct or steal a model) and inference attacks (deducing sensitive information from model outputs). See more on this: Anderson, H; Fordyce A.; Oprea, A; Vassilev, A., *Adversarial Machine Learning: A Taxonomy and Terminology of Attacks and Mitigations*, January 2024, National Institute of Standards and Technology (NIST), p. 18, <https://doi.org/10.6028/NIST.AI.100-2e2023>, 13 April, 2025.

¹³ The Global Risks Report 2024 by the World Economic Forum indicates that disinformation and fake content, fueled by advanced technologies such as generative artificial intelligence, represent the most serious short-term threat to the world. The ability of current AI models to realistically generate content can be exploited to manipulate electoral processes, stock markets, and public discourse. As the report notes, these phenomena may lead to increased social polarization, a decline in public trust, and the destabilization of democratic institutional systems. Additionally, the report identifies threats posed by AI to personal rights — technologies such as deepfakes or voice cloning are already being used to produce pornographic content without the consent of the individuals depicted. See World Economic Forum, *The Global Risks Report 2024*, January 2024, <https://www.weforum.org/publications/global-risks-report-2024/>, 13 April, 2025.

¹⁴ Ajilore, O.; Choi, J. Y.; Kadkol, S.; Kim, J. J. H.; Lee, S.; Nahass, G. R.; Nyugen, T.; Soh, J.; Solomon, I.; Srivatsa, A.V.; Yeh, H., *AI Anxiety: a comprehensive analysis of psychological factors and interventions*, March 2025, AI and Ethics, p. 5-6, <https://doi.org/10.1007/s43681-025-00686-9>, 13 April, 2025.

- lack of comprehensive AI regulations – AI has long operated in a largely unregulated space or under general laws that do not address its unique characteristics, leading to gaps in user protection;
 - unclear legal status of AI – The autonomous nature of AI raises questions about liability and accountability for its decisions and potential damages;
 - mass data processing without adequate oversight – AI systems process vast amounts of sensitive data, increasing risks of unauthorized data collection, user profiling, and privacy violations without proper regulatory safeguards.
4. Ethical Risks – These concern AI's impact on fundamental social values and moral dilemmas:
- dependence on training data quality – AI learns from available datasets, meaning that if those datasets contain biases, the AI will replicate and amplify them, potentially leading to discrimination against specific social groups;
 - scope of AI autonomy – AI's ability to operate independently, especially in critical fields such as justice and healthcare, risks dehumanizing decision-making processes, where individuals become mere objects of algorithmic calculations rather than subjects of ethical, social, and emotional considerations.¹⁵

It is crucial to emphasize that the risks associated with AI are multifaceted and dynamic, and their scope is not exhaustive. The above classification into four main risk categories represents just one possible framework, adopted for the purposes of this discussion to better illustrate the scale and nature of the challenges that AI legislation must address.

¹⁵ Douglas, D. M.; Lacey, J.; Howard, D., *Ethical risk for AI*, August 2024, Springer, p. 1-3, <https://doi.org/10.1007/s43681-024-00549-9>, 13 April, 2025. Steimers, A.; Schneider M., *Sources of Risk of AI Systems*, March 2022, International Journal of Environmental Research and Public Health, p. 18-19, 21-22, <https://www.mdpi.com/1660-4601/19/6/3641>, 13 April, 2025.

4. THE ANTHROPOCENTRIC PARADIGM OF AI REGULATION

4.1. The EU approach to AI regulation: a brief overview

The regulation of AI has been integrated into the EU's *Europe fit for the Digital Age strategy*,¹⁶ which has been one of the key priorities of the European Commission for the years 2019–2024. This process was gradual and multi-stage, preceded by extensive expert analyses, discussions, and public consultations officially launched in 2018.¹⁷ Recognizing both the immense potential and risks associated with AI development, EU institutions took the first concrete steps toward establishing a coherent and unified regulatory policy. Below are the key stages of this process that had a decisive impact on the final shape of AI regulations within the EU.

The starting point for the EU's further strategies on AI was the *Declaration of Cooperation on Artificial Intelligence*, signed in April 2018, which set the direction for a coordinated approach among member states towards this technology.¹⁸ Based on the principles outlined in the declaration, the European Commission published the *Artificial Intelligence for Europe* communication later that same month. This document served as a clear political signal that the EU not only recognized the growing significance of AI but also intended to actively shape its development and regulation. It was also an official confirmation that AI would become one of the pillars of the EU's digital strategy.

In the document, the European Commission outlined recommendations for establishing a robust European strategic framework, emphasizing that the EU must

¹⁶ It is worth mentioning, however, that although the EU's key efforts in shaping AI regulation took place in the early second decade of the 21st century, its interest in this area dates back to 2012. That year marked the beginning of the first initiatives aimed at developing a unified EU strategy for emerging technologies, including artificial intelligence. One significant research initiative from this period was the RoboLaw project (*Regulating Emerging Robotic Technologies in Europe: Robotics facing Law and Ethics*), conducted between 2012 and 2014. The project aimed to analyze the impact of emerging technologies on legal norms, social values, and ethics, as well as assess the adequacy of existing legal frameworks. See The Community Research and Development Information Service (CORDIS), *Project RoboLaw* (ID: 289092), <https://cordis.europa.eu/project/id/289092/reporting/en>, 13 April, 2025.

¹⁷ See the European Commission press release, *Artificial Intelligence: Commission Kicks Off Work on Marrying Cutting-Edge Technology and Ethical Standards*, March 2018, https://ec.europa.eu/commission/presscorner/detail/en/ip_18_1381, 13 April, 2025.

¹⁸ See the European Commission press release, *EU Member States sign up to cooperate on artificial intelligence*, April 2018, <https://digital-strategy.ec.europa.eu/en/news/eu-member-states-sign-cooperate-artificial-intelligence>, 13 April, 2025.

strengthen its technological and industrial capabilities while ensuring balanced development among member states in the digital transition. A key objective was also the need to increase investment in AI research and development, enabling Europe to compete with global leaders in this field, such as the US and China.¹⁹ At the same time, the document emphasized that the development of AI cannot take place in isolation from the EU's fundamental values and fundamental rights.²⁰ The Commission advocated for a balanced approach", combining the benefits of technological advancement with the need to protect individuals and safeguard the interests of society. Accordingly, it announced the necessity of establishing a coherent regulatory framework that addresses both legal and ethical aspects to ensure the safe and responsible deployment of AI.²¹

Another significant milestone in shaping the EU's AI regulations was the publication of the *White Paper on Artificial Intelligence: A European Approach to Excellence and Trust* (hereinafter White Paper)²² in February 2020. This document summarized previous initiatives and outlined future regulatory steps, reaffirming the necessity of maintaining a balanced approach to AI development. The vision presented in the White Paper was built on two key pillars: the ecosystem of excellence and the ecosystem of trust. The first pillar focused on fostering AI research, innovation, and investment, particularly through collaboration between Member States and the private sector. The second pillar aimed to ensure that AI development in Europe aligns with the EU's fundamental values and does not threaten citizens' fundamental rights.²³

As part of these principles, the European Commission recommended the Trustworthy AI concept, which must meet the following criteria: human oversight, technical robustness and safety, privacy and data governance, transparency, diversity, non-discrimination and fairness, social and environmental well-being, and accountability.²⁴ It was emphasized that adherence to these principles is

¹⁹ Communication Artificial Intelligence for Europe, p. 2-3.

²⁰ For values and fundamental rights that, in the EU's view, are particularly at risk in the case of unregulated AI use, see section 4.3. of this article.

²¹ Communication Artificial Intelligence for Europe, p. 3-4.

²² The European Commission, White Paper on Artificial Intelligence - A European approach to excellence and trust, COM(2020) 65 final, Brussels, 19 February 2020, <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex:52020DC0065>, 13 April, 2025.

²³ White Paper, p. 3, 6-29.

²⁴ These guidelines were developed by the High-Level Expert Group on AI, established by the European Commission in 2018. For more information, see Directorate-General for Communications Networks, Content and Technology (European Commission) and High-Level Expert Group on AI, Ethics Guidelines for Trustworthy AI, Publications Office, 2019, p. 11, <https://data.europa.eu/doi/10.2759/346720>, 13 April, 2025.

crucial for increasing public trust in AI, and their implementation should be a prerequisite for the deployment of AI systems in the European single market.²⁵ The White Paper also undertook a broader identification of risks associated with the use of AI, highlighting potential material harms (to health, safety, and property) as well as non-material harms (privacy violations, discrimination, and restrictions on civil liberties).²⁶ Based on this assessment, a preliminary risk-based regulatory model was proposed.²⁷

This model was further refined in the first official draft of the regulation establishing harmonized rules on AI,²⁸ presented by the European Commission in April 2021. The proposed regulatory framework was based on the principles outlined in the White Paper and the results of public consultations.²⁹ In the draft, the Commission set out four main objectives:

- AI system safety – ensuring that the deployment of AI does not violate fundamental rights or EU values,
- Effective enforcement of fundamental rights protection – implementing control mechanisms and accountability measures to prevent discrimination, privacy violations, and non-transparent algorithmic decision-making,
- Legal certainty for investment and innovation – creating a stable legal framework that remains resilient to the rapid evolution of AI technologies, reducing the need for frequent legislative amendments,
- A unified AI market in the EU – preventing regulatory fragmentation and enabling businesses to operate across the Union without the need to comply with divergent national regulations.

One of the key mechanisms to achieve these objectives was the risk-based AI regulatory model. This approach tailors regulatory requirements to the level of risk an AI system may pose to users and society.

²⁵ White Paper, p. 11.

²⁶ See more on this topic in section 4.2. of this article.

²⁷ White Book, p. 19-21.

²⁸ The European Commission, Proposal for a regulation of the European Parliament and of the Council laying down harmonised rules on Artificial Intelligence (Artificial Intelligence Act) and amending certain Union legislative acts, COM(2021) 206 final, Brussels, 21 April 2021, <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:52021PC0206>, 13 April, 2025.

²⁹ See the European Commission press release, *White Paper on Artificial Intelligence: a Public Consultation Towards a European Approach to Excellence and Trust*, February 2020, <https://digital-strategy.ec.europa.eu/en/library/white-paper-artificial-intelligence-public-consultation-towards-european-approach-excellence-and>, 13 April, 2025.

The draft Regulation, following the ordinary legislative procedure, underwent further legislative work. Throughout the process, numerous EU bodies and institutions submitted their positions, opinions, and proposed amendments, including the European Economic and Social Committee,³⁰ the Council of the EU,³¹ the European Central Bank,³² the Committee of the Regions,³³ and the European Parliament.³⁴ As a result of these deliberations, nearly three years after the publication of the first AI Act draft, the European Commission announced that a political agreement had been reached on the final shape of AI regulations.³⁵

The AI Act was adopted on June 13, 2024. In its core provisions, the Regulation first defines its purpose,³⁶ subject matter, and scope of application, covering aspects such as the development, market placement, deployment, and use of AI systems. Additionally, it establishes fundamental definitions used in the act, including terms such as "artificial intelligence system", "provider", and "user" (Chapter I).

³⁰ See more on this in the European Economic and Social Committee, Opinion of the European Economic and Social Committee on Proposal for a Regulation of the European Parliament and of the Council laying down harmonised rules on artificial intelligence (Artificial Intelligence Act) and amending certain union legislative acts, [COM(2021) 206 final — 2021/106 (COD)], EESC 2021/02482, OJ C 517 of 22/12/2021, p. 61–66.

³¹ See more on this in the Council of the European Union, Proposal for a Regulation of the European Parliament and of the Council laying down harmonised rules on artificial intelligence (Artificial Intelligence Act) and amending certain Union legislative acts - General approach, 2021/0106(COD), Brussels, 25 November 2021, <https://data.consilium.europa.eu/doc/document/ST-14954-2022-INIT/en/pdf>, 13 April, 2025.

³² See more on this in the European Central Bank, Opinion of the European Central Bank of 29 December 2021 on a proposal for a regulation laying down harmonised rules on artificial intelligence, (CON/2021/40) 2022/C 115/05, CON/2021/40, OJ C 115 of 11/03/2022, p. 5–11.

³³ See more on this in the European Committee of the Regions, Opinion of the European Committee of the Regions — European approach to artificial intelligence — Artificial Intelligence Act (revised opinion), COR 2021/02682, OJ C 97 of 28/02/2022, p. 60–85.

³⁴ See more on this in the European Parliament, European Parliament Legislative Resolution of 14 June 2023 on the proposal for a regulation of the European Parliament and of the Council laying down harmonised rules on artificial intelligence (Artificial Intelligence Act) and amending certain Union legislative acts, P9_TA(2023)0236, https://www.europarl.europa.eu/doceo/document/TA-9-2023-0236_EN.html, 13 April, 2025.

³⁵ See the European Commission press release, Commission welcomes political agreement on Artificial Intelligence Act, December 2023, https://ec.europa.eu/commission/presscorner/detail/en/ip_23_6473, 13 April, 2025.

³⁶ Art. 1 item (1) of AI Act: "The purpose of this Regulation is to improve the functioning of the internal market and promote the uptake of human-centric and trustworthy artificial intelligence (AI), while ensuring a high level of protection of health, safety, fundamental rights enshrined in the Charter, including democracy, the rule of law and environmental protection, against the harmful effects of AI systems in the Union and supporting innovation".

Subsequently, the Regulation establishes a catalog of AI-related practices deemed unacceptable (Chapter II). It also defines the criteria for classifying high-risk AI systems, the rules governing their use, and the obligations imposed on providers and entities deploying such systems. Additionally, it introduces provisions regarding notifying authorities and notified bodies, outlining their role in the conformity assessment process for AI systems to ensure compliance with regulatory requirements (Chapter III).

In the following sections, the Regulation addresses transparency obligations for providers and entities deploying certain AI systems (Chapter IV) and regulates general-purpose AI models. It defines classification rules and requirements for their market placement, including safety and transparency obligations. Additionally, it provides for the use of codes of practice as a tool to help providers fulfill their obligations effectively (Chapter V).

Furthermore, the regulation includes measures to support innovation, particularly by promoting AI research and development and facilitating small and medium-sized enterprises (SMEs). It also establishes provisions for regulatory sandboxes, allowing for the testing of new AI technologies in controlled environments (Chapter VI).

Next, the AI Act addresses the governance of AI systems within the EU. It defines the supervisory structure and the roles of institutions responsible for implementing and enforcing the regulations (Chapter VII). Additionally, it establishes an EU database for high-risk AI systems, aimed at monitoring and registering information about these systems (Chapter VIII).

The Regulation also includes provisions on post-market monitoring, information exchange, and market surveillance. It outlines the obligations of providers to report and disclose potential risks associated with AI systems (Chapter IX). Additionally, it introduces codes of conduct and guidelines for AI users, aiming to facilitate compliance with regulations and promote best practices in AI governance (Chapter X).

The final section of the AI Act addresses the delegation of powers and committee procedures, establishing mechanisms to adapt the Regulation to future technological developments (Chapter XI). It defines a penalty system for non-compliance, outlining the consequences for entities failing to meet their obligations in the AI sector (Chapter XII). Lastly, it includes final provisions,

specifying the dates for the phased implementation of different chapters³⁷ (Chapter XIII).

4.2. Key Legal Mechanism in the AI Act: A General Framework

In simple terms, the protection of EU values and fundamental rights under the AI Act is based on a two-track regulatory model, which aims, on the one hand, to eliminate the risk of their violation and, on the other, to promote the development of Trustworthy AI.

The central element of this model is a risk-based approach that classifies AI systems according to the level of threat they pose to protected values and rights. Within this framework, key areas most vulnerable to the negative effects of AI use are identified, followed by the determination of appropriate regulatory measures and requirements proportional to the potential risk posed by a given system.

During the design phase of the regulatory framework, AI systems were generally divided into three categories: (a) unacceptable risk, (b) high risk, and (c) low or minimal risk. However, it is worth noting that the final version of the regulation explicitly defines only the first two categories. There is no direct reference to low- or minimal-risk systems, which raises certain interpretative uncertainties.

The first category includes AI systems that, when used in the manner specified in Article 5 of the AI Act, directly threaten protected values. In this context, unacceptable risk refers to dangers arising from the improper use of these technologies. These include:

- systems that manipulate user behavior through subliminal, manipulative, or deceptive techniques that limit their ability to make informed decisions,
- systems that exploit the vulnerabilities of sensitive individuals, such as children, the elderly, or people with disabilities, to influence their decisions in a way that may harm them,
- social scoring systems that assess individuals based on their social behavior, personal traits, or personality characteristics, potentially leading to unfair or discriminatory treatment,

³⁷ The AI Act entered into force on August 1, 2024, while the first provisions concerning prohibited AI practices came into effect on February 2, 2025. Further regulations, including those related to notifying authorities, general-purpose AI models, and penalties, will come into force on August 2, 2025. Full application of the provisions regarding high-risk AI systems will take effect on August 2, 2026. See more on this in Article 113 of the AI Act.

- crime prediction systems that evaluate the likelihood of committing an offense solely based on individual profiling,
- uncontrolled biometric data collection systems, such as the mass gathering of facial images from the internet or surveillance footage to create databases,
- emotion analysis systems in professional and educational contexts that monitor and interpret the emotions of employees and students,
- biometric categorization systems that classify people based on their biological traits, such as race, political beliefs, or sexual orientation,
- remote real-time biometric identification systems in public spaces, such as facial recognition technologies used directly by law enforcement authorities.

The use of these systems for the purposes mentioned above is prohibited under the Regulation.

The second category consists of high-risk AI systems. These include systems that meet both of the following criteria: (a) they are intended for use as a safety-related component of a product covered by EU harmonized legislation, listed in Annex I of the AI Act, or they themselves constitute such a product (Article 6 item 1a AI Act) and (b) they are subject to mandatory third-party conformity assessment before being placed on the market or put into service (Article 6 item 1b AI Act).

This primarily applies to AI systems used in areas where maintaining high safety standards is essential, including aviation, railway transport, road vehicles, elevators, cableways, industrial machinery, and pressure equipment.

High-risk AI systems also include those listed in Annex III of the AI Act. For example, this applies to systems used in areas such as critical infrastructure, education and vocational training, employment, workforce management and access to self-employment, as well as the administration of justice and democratic processes.

The marketing and deployment of high-risk AI systems are permitted provided that they comply with the obligations set out in Articles 8–15 of the AI Act. These include ensuring high-quality datasets used to train the AI system (Article 10 AI Act), developing and maintaining documentation necessary for assessing the system's compliance and intended purpose (Article 11 AI Act), fulfilling information obligations towards entities using AI systems while ensuring an adequate level of transparency (Article 13 AI Act), and guaranteeing human oversight of the system (Article 14 AI Act).

With regard to low- or minimal-risk AI systems, although the Regulation does not explicitly regulate them, it should be noted that they are subject to general transparency obligations outlined in Chapter IV. These include the requirement to disclose when a person is interacting with an AI system (Article 50 item 1 AI Act) and the obligation to inform about the artificial origin of generated content and ensure its proper labeling (Article 50 item 2 AI Act).

4.3. The human-centric approach – to protect values and fundamental rights

The anthropocentric paradigm of AI regulation in EU law finds its roots in the European values enshrined in Article 2³⁸ of the Treaty on European Union, and the rights guaranteed by the Charter of Fundamental Rights of the European Union (hereinafter: CFR). In the context of the risks identified in Section 3 of this article, it is evident that the uncontrolled use of AI systems may seriously endanger the realisation of those values and lead to violations of individual rights.³⁹

An analysis of the directions outlined in EU policy and legal documents allows for the formulation of specific guidelines that shape the EU legislator's approach to regulating artificial intelligence in relation to protected rights:

³⁸ "The Union is founded on the values of respect for human dignity, freedom, democracy, equality, the rule of law and respect for human rights, including the rights of persons belonging to minorities. These values are common to the Member States in a society in which pluralism, non-discrimination, tolerance, justice, solidarity and equality between women and men prevail."

³⁹ See, e.g., White Paper, p. 14–15. The EU has identified threats to its values and fundamental rights in numerous documents. For example, in the report by the Committee of Experts on Internet Intermediaries, Algorithms and Human Rights – Study on the Human Rights Dimensions of Automated Data Processing Techniques and Possible Regulatory Implications, it is noted that the development of algorithms and artificial intelligence poses risks to the right to a fair trial, privacy protection, freedom of expression, and may lead to discrimination and electoral manipulation. Algorithms used in the judicial system may reinforce biases and reduce decision-making transparency, while predictive policing technologies employed in law enforcement databases may disproportionately affect minority groups. Furthermore, user profiling in social media and advertising systems impacts media pluralism, creating so-called filter bubbles and increasing the risk of disinformation. See more on this the Committee of experts on internet intermediaries (Council of Europe), Algorithms and human rights - Study on the human rights dimensions of automated data processing techniques and possible regulatory implications, March 2018, <https://edoc.coe.int/en/internet/7589-algorithms-and-human-rights-study-on-the-human-rights-dimensions-of-automated-data-processing-techniques-and-possible-regulatory-implications.html#>, 15 March, 2025. For more, see e.g. the Council of the EU, Presidency Conclusions on the Charter of Fundamental Rights in the Context of AI and Digital Transformation, 11481/20, Brussels, 21 October 2020, <https://data.consilium.europa.eu/doc/document/ST-11481-2020-INIT/pl/pdf>, 13 April, 2025.

- human dignity (Article 1 CFR) – AI development should always uphold human primacy, avoiding situations where algorithms replace individual autonomous decisions or lead to the dehumanization of individuals,
- respect for private and family life (Article 7 CFR) and protection of personal data (Article 8 CFR) – AI technologies must comply with data protection regulations, ensuring users retain control over their personal information and preventing excessive intrusion into private life,
- freedom of expression and information (Article 11 CFR) and freedom of assembly and of association (Article 12 CFR) – AI use in media and public spaces must not restrict access to information, result in censorship, or manipulate public opinion,
- non-discrimination (Article 21 CFR) and equality between women and men (Article 23 CFR) – AI must not reinforce biases or lead to unequal treatment based on gender, age, origin, or other personal characteristics,
- the rights of the child (Article 24 CFR) and the integration of persons with disabilities (Article 26 CFR) – AI development should consider vulnerable groups, ensuring equal access to services and protection from the potential negative effects of automation,
- fair and just working conditions (Article 31 CFR) and consumer protection (Article 38 CFR) – AI should not worsen employment conditions or violate consumer rights through unfair business practices or lack of transparency in automated decision-making systems,
- environmental protection (Article 37 CFR) – AI development should follow sustainable principles, taking ecological and public health aspects into account,
- the right to an effective remedy and to a fair trial (Article 47 CFR) and the presumption of innocence and right of defence (Article 48 CFR) – the use of AI in justice systems must ensure transparency, impartiality, the right to defense, and the ability to appeal decisions influenced by algorithmic processes.

The EU's approach is characterised by the belief that the development of artificial intelligence is not an end in itself and must not come at the expense of the established legal, social, and axiological order. The human being remains the key point of reference – not the market or technology – and must be the main beneficiary of the implemented solutions.

5. CONCLUSION

The dynamic development of AI, as well as the growing risks associated with its use, make its regulation an undeniable necessity. While AI holds immense potential for innovation and economic growth, it also comes with significant threats. The absence of appropriate regulations could lead to a situation where this technology spirals out of control, becoming a tool for violating individual rights, deepening social inequalities, or spreading disinformation.

In this context, the EU's approach deserves particular recognition, as it was the first to undertake the creation of a comprehensive legal framework for AI. The anthropocentric paradigm of AI regulation adopted by the EU is based on the premise that technological development, even if it may contribute to the transformation of society, should not come at the expense of fundamental values and rights. The EU model balances technological progress with the need to protect individual rights, democracy, and the rule of law. Consequently, the AI Act is not only a protective instrument but also a tool for building public trust in AI, which is a crucial condition for its continued development in a responsible and socially acceptable manner.

The introduced risk-based approach allows for a tiered level of regulation - more stringent requirements, and even outright bans, are imposed on AI systems that may pose real threats, while low- or minimal-risk systems are subject to fewer restrictions. This approach does not constitute a barrier to innovation but aims to establish a framework for the sustainable and safe deployment of AI.

However, it should be remembered that this is only the beginning of the EU's efforts in AI regulation. The regulation has not yet come into full force, and member states are working on adapting their legal frameworks to the new provisions. The effectiveness of these regulations and their impact on the development and implementation of AI thus remain an open subject for further research.

REFERENCES

BOOKS AND ARTICLES

1. McCarthy, J.; Minsky, M. L.; Rochester, N.; Shannon, C. E., *A Proposal for the Dartmouth Summer Research Project on Artificial Intelligence*, August 1955, <http://jmc.stanford.edu/articles/dartmouth/dartmouth.pdf>, 13 April, 2025.
2. Slattery, P.; Saeri, A. K.; Grundy, E. A. C.; Graham, J.; Noetel, M.; Uuk, R.; Dao, J.; Pour, S.; Casper, S.; Thompson, N., *The AI Risk Repository: A Comprehensive Meta-Review, Database, and Taxonomy of Risks From Artificial Intelligence*, August 2024, <http://dx.doi.org/10.13140/RG.2.2.28850.00968>, 13 April, 2025.
3. Douglas, D. M.; Lacey, J.; Howard, D., *Ethical risk for AI*, August 2024, Springer, <https://doi.org/10.1007/s43681-024-00549-9>, 13 April, 2025.
4. Steimers, A.; Schneider M., *Sources of Risk of AI Systems*, March 2022, International Journal of Environmental Research and Public Health, <https://www.mdpi.com/1660-4601/19/6/3641>, 13 April, 2025.
5. McCarthy, J., *What is Artificial Intelligence?*, November 2007, <https://www-formal.stanford.edu/jmc/whatisai.pdf>, 13 April, 2025.
6. Zednik, C., *Solving the Black Box Problem: A Normative Framework for Explainable Artificial Intelligence*, March 2019, Philosophy & Technology, <https://doi.org/10.48550/arXiv.1903.04361>, 13 April, 2025.
7. Athaluri, S.; Dave, T.; Duddumpudi, R.; Manthena; S.; Kesapragada, V.; Yarlagadda V., *Exploring the Boundaries of Reality: Investigating the Phenomenon of Artificial Intelligence Hallucination in Scientific Writing Through ChatGPT References*, April 2023, Cureus, <https://doi.org/10.7759/cureus.37432>, 13 April, 2025.
8. Anderson, H; Fordyce A.; Oprea, A; Vassilev, A., *Adversarial Machine Learning: A Taxonomy and Terminology of Attacks and Mitigations*, January 2024, National Institute of Standards and Technology (NIST), <https://doi.org/10.6028/NIST.AI.100-2e2023>, 13 April, 2025.
9. Ajilore, O.; Choi, J. Y.; Kadkol, S.; Kim, J. J. H.; Lee, S.; Nahass, G. R.; Nyugen, T.; Soh, J.; Solomon, I.; Srivatsa, A.V.; Yeh, H., *AI Anxiety: a comprehensive analysis of psychological factors and interventions*, March 2025, AI and Ethics, <https://doi.org/10.1007/s43681-025-00686-9>, 13 April, 2025.
10. Bareis, J; Folberth, A.; Jahnel, J.; Orwat, C.; Wadephul, C., *Normative Challenges of Risk Regulation of Artificial Intelligence and Automated Decision-Making*, November 2022, <https://doi.org/10.48550/arXiv.2211.06203>, 13 April, 2025.

LEGAL ACTS

1. Charter of Fundamental Rights of the European Union, OJ C 202/389 of 07/06/2016.
2. Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonised rules on artificial intelligence and amending Regulations (EC) No 300/2008, (EU) No 167/2013, (EU) No 168/2013, (EU) 2018/858, (EU) 2018/1139, and (EU) 2019/2144, as well as Directives 2014/90/EU, (EU) 2016/797, and (EU) 2020/1828, OJ L 2024/1689 of 12/7/2024.
3. Treaty on European Union, OJ C 202/1 of 07/06/2016.

NETWORK RESOURCES

1. OECD Recommendation of the Council on Artificial Intelligence, 22 May 2019, <https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0449>, 13 April, 2025.
2. Oxford English Dictionary, https://www.oed.com/dictionary/intelligence_n, 13 April, 2025.
3. The Committee of experts on internet intermediaries (Council of Europe), Algorithms and human rights - Study on the human rights dimensions of automated data processing techniques and possible regulatory implications, March 2018, <https://edoc.coe.int/en/internet/7589-algorithms-and-human-rights-study-on-the-human-rights-dimensions-of-automated-data-processing-techniques-and-possible-regulatory-implications.html#>, 13 April, 2025.
4. The Community Research and Development Information Service (CORDIS), Project RoboLaw (ID: 289092), <https://cordis.europa.eu/project/id/289092/reporting/en>, 13 April, 2025.
5. The Council of the EU, Presidency Conclusions on the Charter of Fundamental Rights in the Context of AI and Digital Transformation, 11481/20, Brussels, 21 October 2020, <https://data.consilium.europa.eu/doc/document/ST-11481-2020-INIT/pl/pdf>, 13 April, 2025.
6. The Council of the European Union, Proposal for a Regulation of the European Parliament and of the Council laying down harmonised rules on artificial intelligence (Artificial Intelligence Act) and amending certain Union legislative acts - General approach, 2021/0106(COD), Brussels, 25 November 2021, <https://data.consilium.europa.eu/doc/document/ST-14954-2022-INIT/en/pdf>, 13 April, 2025.

7. The Directorate-General for Communications Networks, Content and Technology (European Commission) and High-Level Expert Group on AI, Ethics Guidelines for Trustworthy AI, Publications Office, 2019, <https://data.europa.eu/doi/10.2759/346720>, 13 April, 2025.
8. The European Central Bank, Opinion of the European Central Bank of 29 December 2021 on a proposal for a regulation laying down harmonised rules on artificial intelligence, (CON/2021/40) 2022/C 115/05, CON/2021/40, OJ C 115 of 11/03/2022.
9. The European Commission press release, Artificial Intelligence: Commission Kicks Off Work on Marrying Cutting-Edge Technology and Ethical Standards, March 2018, https://ec.europa.eu/commission/presscorner/detail/en/ip_18_1381, 13 April, 2025.
10. The European Commission press release, Commission welcomes political agreement on Artificial Intelligence Act, December 2023, https://ec.europa.eu/commission/presscorner/detail/en/ip_23_6473, 13 April, 2025.
11. The European Commission press release, EU Member States sign up to cooperate on artificial intelligence, April 2018, <https://digital-strategy.ec.europa.eu/en/news/eu-member-states-sign-cooperate-artificial-intelligence>, 13 April, 2025.
12. The European Commission press release, *White Paper on Artificial Intelligence: a Public Consultation Towards a European Approach to Excellence and Trust*, February 2020, <https://digital-strategy.ec.europa.eu/en/library/white-paper-artificial-intelligence-public-consultation-towards-european-approach-excellence-and>, 13 April, 2025.
13. The European Commission, Communication from the Commission to the European Parliament, the European Council, the Council, the European Economic and Social Committee and the Committee for the Regions, Artificial Intelligence for Europe, COM(2018) 237 final, Brussels, 25 April 2018, <https://eur-lex.europa.eu/legal-content/PL/TXT/PDF/?uri=CELEX:52018DC0237>, 13 April, 2025.
14. The European Commission, Proposal for a regulation of the European Parliament and of the Council laying down harmonised rules on Artificial Intelligence (Artificial Intelligence Act) and amending certain Union legislative acts, COM(2021) 206 final, Brussels, 21 April 2021, <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:52021PC0206>, 13 April, 2025.
15. The European Commission, White Paper on Artificial Intelligence - A European approach to excellence and trust, COM(2020) 65 final, Brussels, 19 February 2020, <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex:52020DC0065>, 13 April, 2025.

16. The European Committee of the Regions, Opinion of the European Committee of the Regions — European approach to artificial intelligence — Artificial Intelligence Act (revised opinion), COR 2021/02682, OJ C 97 of 28/02/2022, p. 60–85.
17. The European Economic and Social Committee, Opinion of the European Economic and Social Committee on Proposal for a Regulation of the European Parliament and of the Council laying down harmonised rules on artificial intelligence (Artificial Intelligence Act) and amending certain union legislative acts, [COM(2021) 206 final — 2021/106 (COD)], EESC 2021/02482, OJ C 517 of 22/12/2021, p. 61–66.
18. The European Parliament, European Parliament Legislative Resolution of 14 June 2023 on the proposal for a regulation of the European Parliament and of the Council laying down harmonised rules on artificial intelligence (Artificial Intelligence Act) and amending certain Union legislative acts, P9_TA(2023)0236, https://www.europarl.europa.eu/doceo/document/TA-9-2023-0236_EN.html, 13 April, 2025.
19. The High-Level Expert Group (European Commission), Artificial Intelligence, Ethics Guidelines for Trustworthy AI, April 2019, https://ec.europa.eu/newsroom/dac/document.cfm?doc_id=60651, 13 April, 2025.
20. University of Cambridge, The Best or Worst Thing to Happen to Humanity? Stephen Hawking Launches Centre for the Future of Intelligence, <https://www.cam.ac.uk/research/news/the-best-or-worst-thing-to-happen-to-humanity-stephen-hawking-launches-centre-for-the-future-of->, 13 April, 2025.
21. World Economic Forum, The Global Risks Report 2024, January 2024, <https://www.weforum.org/publications/global-risks-report-2024/>, 13 April, 2025.