Effects of the Artificial Intelligence Ethical Regulations on Society

Natalia Marinova[1]

Abstract

The development of Artificial Intelligence Systems has advanced remarkably in the last decade and has had a real impact on people, institutions and culture. The current research aspiration to create Artificial General Intelligence Systems can result in both enormous potential benefits (finding a cure for all known diseases, ending poverty, outstanding scientific achievements, etc.) and fatal consequences (ending the human race, military operations with armed machines, sufficient power to destroy the planet) for the human race. Properly addressing these risks requires timely adaptation of Artificial Intelligence legislation to the rapid development of the technology. This paper systematizes key international and national initiatives, policies, and acts for developing ethically responsible machine algorithms and artificial intelligence systems, as well as examines the effects of the Artificial Intelligence Act on society.

Introduction

Advances in the field of artificial intelligence have a broad impact on various sectors of the economy, redefining the way business organizations function (mainly in the direction of transforming jobs, professions and the workforce) and people's engagement with newly emerging digital technologies. The technology of current narrow-purpose artificial intelligence systems has the potential to create fruitful innovations in every sphere of human life, but at the same time, it hides huge risks of a social, moral, economic and even existential nature - algorithmic bias, spreading disinformation, putting thousands of jobs at risk, erasing biological units by super-intelligent machines with artificial intelligence, etc.

Awareness, knowledge and addressing of the different types of risks of the functioning of artificial intelligence systems is a successful prerequisite for the development of future artificial intelligence solutions that do not threaten digital privacy, information security and the right to freedom and autonomy of individuals. Over the past few years, calls for big tech companies to apply ethical principles in the design and creation of artificial intelligence technologies that support and do not harm human development, reflected in open declarations and scientific publications on the subject have become more and more numerous.

Artificial Intelligence Ethical Regulations Acts

The ethical responsibility to conduct research promoting the positive and avoiding or mitigating the expected and unforeseen negative side effects of the created machine algorithms and systems with artificial intelligence with narrow purpose is regulated by several international initiatives, policies and legislative acts. The most significant of them are systematized in chronological order in Table 1 below:

---

[1] Associate Professor, PhD (Econ). Department of Business Informatics, D. A. Tsenov Academy of Economics, Svishtov, ORCID ID: 0000-0001-8732-7564, *e-mail: n.marinova@uni-svishtov.bg*

Table 1: Key International Acts Governing the Ethical Development of AI Systems

Ethical Regulation Acts	Description
Guidelines for AI Procurement (World Economic Forum, 2019)	World Economic Forum recommendations for national procurement in the field of artificial intelligence.
Rome Call (RenAIssance Foundation, 2020)	An international charter with six ethical principles for making self-explanatory, inclusive, impartial, reproducible and accountable narrow-purpose artificial intelligence systems.
State of Implementation of the OECD AI Principles (OECD, 2021)	International policy framework of the Organisation for Economic Co-operation and Development for the responsible development and use of Artificial Intelligence.
Recommendation on the Ethics of Artificial Intelligence (UNESCO, 2021)	A global instrument proposed by UNESCO to define a holistic, inclusive and multicultural framework of interdependent values, principles and actions to responsibly address the known and unknown impacts of AI technologies on people, societies, ecosystems, and the environment.
Global Partnership on Artificial Intelligence (GPAI Council, 2022)	Global initiative stating the need to develop narrow-purpose artificial intelligence systems in line with human rights and democratic societal values. The fourth session of the Council of the Global Partnership on AI ends with the signing of a declaration by the ministers of the member states reaffirming their commitment to the principles of the Organisation for Economic Co-operation and Development on artificial intelligence.
Executive Order on AI (The White House, 2023)	A set of standards for developing safe, secure, and trustworthy artificial intelligence that protects Americans' privacy, promotes justice and civil rights, protects consumers and workers, promotes innovation and competition, develops American leadership around the world, and more, signed by 15 leading U.S. technology companies.
International Code of Conduct for Organizations Developing Advanced AI Systems (European Commission, 2023)	A global voluntary agreement between G7 leaders and artificial intelligence system manufacturers recommends the behaviour of responsibly developing safe, secure and reliable advanced artificial intelligence systems, large language models and generative AI solutions.
Bletchley Declaration (United Kingdom Government, 2023)	An agreement between the European Union, 27 countries and major technology companies to work together to test the safety of their new artificial intelligence products before they are officially released to the public. While it has no regulatory impact, the declaration promises that signatories will work together on shared safety standards.
Resolution on Artificial Intelligence (United Nations, 2024)	A global resolution on artificial intelligence, proposed by the US and supported by China and over 120 other countries, encourages countries to guarantee human rights, protect personal data and monitor the risks associated with the development of the technology.
Artificial Intelligence Act (European Union, 2024)	A legislative framework with harmonised rules for the development of credible and fundamental human rights Artificial Intelligence that provides developers and deployers of artificial intelligence systems with clear requirements and obligations on the specific uses of the technology.
AI Pact (European Commission, 2024)	A set of voluntary commitments to implement the principles of the European Artificial Intelligence Act signed between the EU AI Office and more than 100 companies, reinforcing their commitment to the development of artificial intelligence management strategies, to mapping high-risk artificial intelligence systems, to increasing the knowledge of artificial intelligence workers, to ensure human oversight, to mitigate risks and transparent labelling of certain types of content generated by AI, etc.

In addition to the international level, numerous national initiatives have been launched in the field of artificial intelligence in more than 69 countries and territories (OECD.AI, 2021). According to a survey of 25 countries with approved regulations in the field of artificial intelligence (Zhang, et al., 2022, p. 176), the largest number of regulatory documents were adopted in the United States, Russia, Belgium, Spain and the United Kingdom. National artificial intelligence strategies have been adopted in more than 30 countries, including Bulgaria, and Bangladesh, Malaysia and Tunisia are in the process of developing them (UNESCO, 2021).

The Bulgarian concept for the development of artificial intelligence until 2030 (Ministry of Transport and Communications, 2020) is in line with the initiatives of the European Commission, which considers artificial intelligence as one of the main drivers of digital transformation in Europe and as a significant factor in ensuring the competitiveness of the European economy. According to an analysis by the Economic and Social Council of the Republic of Bulgaria, "although our country is not among the leaders on the old continent in terms of adaptation and use of artificial intelligence solutions, the implementation of such technologies will increase the country's economic growth" (Economic and Social Council, 2024). The document contains 19 conclusions and recommendations on how to harmonize Bulgarian legislation with the European rules on artificial intelligence in the next two years.

Regulatory Effects of the European Artificial Intelligence Act on Manufacturers and Users of Artificial Intelligence Technologies

Despite the existence of numerous current international and national legislative initiatives for the ethical creation of artificial intelligence systems, the leading regulator in the technology sector is the region of the Old Continent. Over the past few years, a set of legislation has been introduced in Europe aimed at regulating the digital economy in different countries: the General Data Protection Regulation (GDPR) entered into force in 2018 (European Commission, 2018), the Digital Service Act was published in 2020 (European Parliament, 2020), in 2022, the European Data Governance Act (European Commission, 2022), the Digital Markets Act (European Commission, 2022) and the Cyber Resilience Act (European Commission, 2022) was activated, and this year the long-awaited Artificial Intelligence Act officially came into force (European Union, 2024).

The desire of the European Parliament and the Council of the European Union to establish the continent as the first world region to adopt comprehensive regulation in the field of artificial intelligence has a non-territorial effect, as the Artificial Intelligence Act (AI Act) refers not only to entities within the European Union but also to developers, implementers, importers and distributors of artificial intelligence systems outside the European Union if the result of their system is consumed within the Union. The scope of the regulation covers large, medium, small and micro-enterprises from the private and public sectors that develop, market, import into the EU, implement and use such solutions[1] in their activities.

The AI Act will begin to apply 24 months after it enters into force on August 1, 2024, enabling business organizations to proactively start analysing gaps in their artificial intelligence systems, assess their risks, develop their own ethical rules for creating new solutions with such functionality, and train their employees to implement them. At present, many business organizations are not aware of the binding effect of the

---

[1] The definition of an AI system in the AI Act is quite broad to cover the maximum range of current and future developed and used machine-based systems that are designed to operate with different levels of autonomy, which can be adaptable after implementation, which, based on implicit or explicit input data, output results in the form of predictions, content, recommendations or solutions, and which may have an impact on the physical or virtual environment.

European regulatory framework and in order not to lag behind in the projects for its implementation, they will need legal and technical expertise[1] in the field of information technology.

To comply with the AI Act, several European bodies with regulatory and advisory prerogatives have been created (such as the AI Office and an AI Board), and the enforcement of the provisions of the act is imputed to the national public authorities in the Member States, which, without adopting implementing legislation, can detail the rules in the regulation in the direction of their practical application within the specific country. Fines for infringements of the AI Act vary depending on the severity of the infringement, with the European provision setting the limits of penalties[2] and Member States setting the rules on financial penalties and other coercive measures (warnings and non-pecuniary impact measures).

The AI Act follows a risk-based approach and classifies AI systems into four categories: prohibited AI with unacceptable risk, high-risk AI systems, general-purpose AI (GPAI) and foundation models with limited risk, and low-risk AI systems (see Figure 1 below):
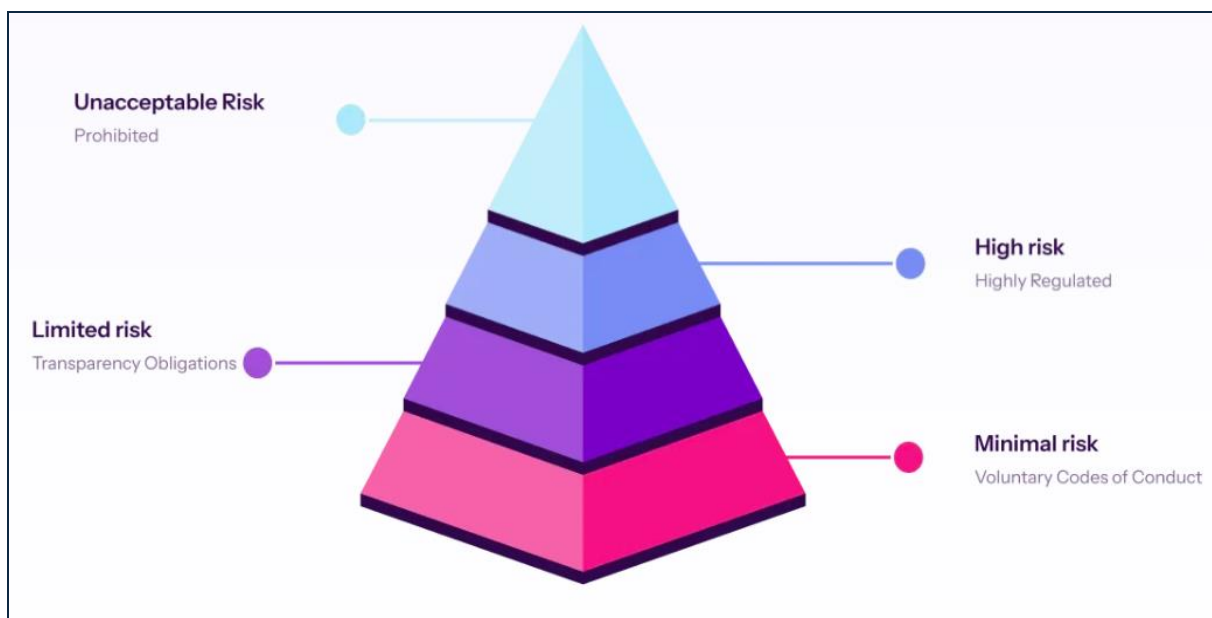
Source: What is the EU AI act?, 2024

Figure 1: Classification of Artificial Intelligence Systems under the Artificial Intelligence Act

---

[1] To help comply with provisions in the AI Act and to alleviate the effects of the introduction of the regulation, the European Commission has envisaged the creation of a regulatory sandbox in the Member States, providing a controlled environment to promote the development, training, testing and validation of innovative artificial intelligence systems. To reduce the administrative burden for small, medium, micro and start-up enterprises, lower fees are envisaged for assessing the compliance of their solutions with the requirements of the Regulation, trainings are provided, etc.

[2] The most significant pecuniary penalties are for non-compliance with the prohibition on placing on the market, putting into operation and use of prohibited artificial intelligence systems - up to EUR 35,000,000 or, if the offender is an enterprise, up to 7% of its total annual worldwide turnover for the previous financial year, whichever is higher. For violation of other provisions, a penalty of up to EUR 15,000,000 is provided or, if the offender is an enterprise, up to 3% of its total annual worldwide turnover for the previous financial year, whichever is higher. If a debtor provides inaccurate, incomplete or misleading information in response to a request from a competent authority, the penalty is up to EUR 7 500 000 or, if the infringer is an enterprise, up to 1% of its total annual worldwide turnover for the previous financial year, whichever is higher.

1) **AI systems with unacceptable risk**. This category includes information and social practices that the legislation seeks to prohibit, the main of which are:

    a. Human manipulation. Prohibited for marketing, commissioning and use are artificial intelligence systems that influence a person's behaviour on a subconscious level (prompting him to make an atypical, irrational or threatening decision), impairing his ability to make an informed decision.

    b. Social Assessment. Artificial intelligence systems that assess or classify a person on the basis of their social behaviour or personality characteristics, that create or extend facial recognition databases through unintentional extraction of facial images from the internet or CCTV recordings, and that assess the emotional state of employees (except where the system is deployed for medical or safety reasons) shall be prohibited.

    c. Unfavourable treatment. The list of prohibited also includes law enforcement systems for remote biometric identification of people in real-time in publicly accessible places (with a few exceptions), which can generate an assessment of their future criminal behaviour.

2) **AI systems with high risk**. According to European regulation, activities that determine the access of individuals to financial resources or basic utilities are subject to strict control, the granting of access to which by an automated system with artificial intelligence may be hindered by the application of a racial, ethnic, gender or other type of existing or new socially discriminatory model. Artificial intelligence systems classified as high-risk will also be able to be used after the AI Act enters into force, if the providers of such solutions meet the criteria in the regulation, and the business organizations using such solutions have trained, competent and empowered employees to implement it. Examples of activities affecting the rights and safety of individuals are:

    a. Personnel selection (specifically referring to the activities of publishing job advertisements, analysing or filtering job applications and evaluating candidates for a given position).

    b. Remote biometric identification of persons.

    c. Recognition of human emotions.

    d. Management of critical transport and utility infrastructure (road, electricity, telecommunications, etc.).

    e. Provision of key educational services and forms of employment.

    f. Provision of basic private and public services (creditworthiness assessment, life insurance risk assessment, emergency call classification, etc.).

3) **AI systems with limited risk**. General-purpose AI and foundation models face specific transparency requirements and the quality of the data they are trained on. The safe, unbiased and non-discriminatory functioning of such artificial intelligence systems requires the implementation of appropriate administration and management practices ensuring the completeness, safety[1], confidentiality[2], appropriateness and representativeness of their training and validation datasets. AI Act obliges GPAI model providers to comply with European Union copyright rules.

---

[1] In order to prevent the processes of manipulation of training data or GPAI models, it is required to implement technical measures to detect, react, neutralize and control cyberattacks to them.

[2] Special categories of personal data may be processed after the introduction of additional GDPR requirements guaranteeing the fundamental rights and freedoms of individuals.

4) **AI systems with minimal risk**. Low-risk artificial intelligence systems need to ambiguously ensure that their users are aware that they are consuming AI-generated content. The main problems in the functioning of today's generative artificial intelligence solutions are related to:

    a. Failure to comply with GDPR rules in whole or in part.

    b. Failure to comply with intellectual property protection legislation.

    c. Violation of confidentiality and trade secret requirements (contractual or legal).

    d. Failure to provide users with sufficient information about how the system operates, risks of use, and information that the user shares or consumes as a result of the operation of the system.

    e. Improper design and loading of the artificial intelligence system with relevant, correct, true and compliant data, which can lead to the generation of inaccurate, discriminatory, misleading, erroneous and unlawful results and advice.

Conclusion

In the last five years, a number of important documents have been created on a global and national scale regulating the development of ethically acting Artificial Intelligence Systems. Currently, the creators of the most complete and stringent regulations in the field of technology are the bodies of the European Union, which in a short period of time have developed and introduced a set of legislative acts to regulate the digital economy of the Old Continent. A serious step in protecting the rights of individuals around the world is the AI Act introduced on August 1, 2024, prohibiting the production, distribution and application of unacceptably high-risk artificial intelligence systems and introducing a wide range of requirements for permitted ones. The European regulation obliges designers and developers of almost all artificial intelligence solutions designed to interact directly with individuals to respect the ethical rights of consumers for informed and safe use of their products. A positive step in the document is also the introduction of rules for general-purpose AI and foundation models, which are implemented in other artificial intelligence systems.

References

Economic and Social Council. (2024, October 4). *Analysis of the expected effects of the impact of artificial intelligence on the labor market in Bulgaria: Proposals for specific measures to overcome the problems (in Bulgarian).* Retrieved from https://esc.bg/: https://esc.bg/wp-content/uploads/2024/10/ESC_4_068_2024.pdf

European Commission. (2018, May 25). *General data protection regulation.* Retrieved from https://commission.europa.eu/index_en: https://gdpr-info.eu/

European Commission. (2022, September 15). *Cyber resilience act.* Retrieved from https://ec.europa.eu/: https://digital-strategy.ec.europa.eu/en/library/cyber-resilience-act

European Commission. (2022, September 14). *Digital markets act (DMA) legislation.* Retrieved from https://commission.europa.eu/index_en: https://digital-markets-act.ec.europa.eu/legislation_en

European Commission. (2022, May 30). *European data governance act.* Retrieved from https://commission.europa.eu/index_en: https://digital-strategy.ec.europa.eu/en/policies/data-governance-act

European Commission. (2023, October 30). *Hiroshima process international code of conduct for advanced AI systems.* Retrieved from https://commission.europa.eu/index_en: https://digital-strategy.ec.europa.eu/en/library/hiroshima-process-international-code-conduct-advanced-ai-systems

European Commission. (2024, September 25). *AI pact.* Retrieved from https://commission.europa.eu/index_en: https://digital-strategy.ec.europa.eu/en/policies/ai-pact

European Parliament. (2020, October 27). *2020/0361(COD) Digital services act.* Retrieved from https://europarl.europa.eu/: https://oeil.secure.europarl.europa.eu/oeil/popups/ficheprocedure.do?reference=2020/0361(COD)&l=en

European Union. (2024, June 13). *Artificial intelligence act.* Retrieved from https://eur-lex.europa.eu/: https://eur-lex.europa.eu/eli/reg/2024/1689/oj

GPAI Council. (2022, November 22). *GPAI 2022 ministers' declaration.* Retrieved February 24, 2023, from https://www.gpai.ai/: https://www.gpai.ai/events/tokyo-2022/ministerial-declaration/GPAIMinistersDeclaration2022.pdf

Ministry of Transport and Communications. (2020, December 16). *Concept for the development of Artificial Intelligence in Bulgaria until 2030: Artificial inteligence for smart growth and a prosperous democratic society (in Bulgarian).* Retrieved October 16, 2024, from https://www.mtc.government.bg/: https://www.mtc.government.bg/sites/default/files/conceptforthedevelopmentofaiinbulgariauntil2030.pdf

OECD. (2021, June). *State of implementation of the OECD AI principles: Insights from national AI policies.* Retrieved February 25, 2023, from https://www.oecd-ilibrary.org/: ttps://www.oecd-ilibrary.org/science-and-technology/state-of-implementation-of-the-oecd-ai-principles_1cd40c44-en

OECD.AI. (2021). *National AI policies & strategies.* Изтеглено на 25 February 2023 г. от https://oecd.ai/: https://oecd.ai/en/dashboards/overview

RenAIssance Foundation. (2020, February 28). *Rome call for AI ethics.* Retrieved February 25, 2023, from https://www.romecall.org/: https://www.romecall.org/wp-content/uploads/2022/03/RomeCall_Paper_web.pdf

The White House. (2023, October 30). *Fact Sheet: President Biden issues executive order on safe, secure, and trustworthy artificial intelligence.* Retrieved from https://www.whitehouse.gov/: https://www.whitehouse.gov/briefing-room/statements-releases/2023/10/30/fact-sheet-president-biden-issues-executive-order-on-safe-secure-and-trustworthy-artificial-intelligence/

UNESCO. (2021, November 23). *Recommendation on the ethics of artificial intelligence.* Retrieved February 25, 2023, from https://unesco.org/: https://unesdoc.unesco.org/ark:/48223/pf0000381137

UNESCO. (2021). *UNESCO science report: The race against time for smarter development.* (S. Schneegans, T. Straza, & J. Lewis, Ред.) Paris: UNESCO Publishing.

United Kingdom Government. (2023, November 1). *The Bletchley declaration by countries attending the AI safety summit, 1-2 November 2023.* Retrieved from https://www.gov.uk/: https://www.gov.uk/government/publications/ai-safety-summit-2023-the-bletchley-declaration/the-bletchley-declaration-by-countries-attending-the-ai-safety-summit-1-2-november-2023

United Nations. (2024, March 21). *General Assembly adopts landmark resolution on Artificial Intelligence.* Retrieved from https://news.un.org/: https://documents.un.org/doc/undoc/ltd/n24/065/92/pdf/n2406592.pdf

What is the EU AI act? (2024). Изтеглено на 16 October 2024 г. от https://www.credo.ai/: https://www.credo.ai/eu-ai-act?utm_term=the%20ai%20act&utm_campaign=EU+AI+Act&utm_source=bing&utm_medium=ppc&hsa_acc=9234903900&hsa_cam=20678021731&hsa_grp=1328212367439342&hsa_ad=&hsa_src=o&hsa_tgt=kwd-83014371791027:loc-26&hsa_kw=the%20ai%20act&hsa_mt=p

World Economic Forum. (2019). *Guidelines for AI procurement.* Изтеглено на 25 February 2023 г. от https://www3.weforum.org/:
https://www3.weforum.org/docs/WEF_Guidelines_for_AI_Procurement.pdf

Zhang, D., Maslej, N., Brynjolfsson, E., Etchemendy, J., Lyons, T., Manyika, J., . . . Perrault, R. (2022, March). *The AI index 2022 annual report.* Retrieved January 24, 2023, from https://aiindex.stanford.edu/: https://aiindex.stanford.edu/wp-content/uploads/2022/03/2022-AI-Index-Report_Master.pdf