https://doi.org/10.38044/2686-9136-2022-3-3-43-57

ARTICLES

ARTIFICIAL INTELLIGENCE AND HUMAN RIGHTS: WHAT IS THE EU's APPROACH?

Anna Y. Marchenko*, Mark L. Entin

Moscow State Institute of International Relations (MGIMO-University)
76, ave. Vernadsky, Moscow, Russia, 119454

Abstract

Threats posed to human rights by the rapid development of artificial intelligence (AI) are considered, along with some potential legal mitigations. The active efforts of the EU in the field of AI regulation seem particularly relevant for research considering its approach centred on citizens' rights. Thus, the present study aims to describe the key features of the EU approach to regulating AI in the context of human rights protection, as well as identifying both its achievements and deficiencies, and proposing improvements to existing provisions. The presented analysis of the proposed AI Act pays special attention to provisions that set out to eliminate or mitigate the main risks and dangers of AI. The currently intensive development of AI regulation in the EU (the Presidency Compromise Text presented by the Council of the EU, amendments of the European Committee of the Regions, opinions of interested parties and human rights organisations, etc.) makes this study especially timely due to its highlighting of problematic aspects. The analysis shows that, on closer examination, the proposed law leaves many sensitive and controversial issues unsettled. In the context of AI applications, the proposed solution is considered as an emergency measure in order to rapidly integrate purportedly trustworthy AI into human society. As a result of the analysis, the authors propose potential improvements to the AI Act, including the possibility to update the lists of all types of AI, clarify the concept of transparency and eliminate the self-assessment procedure. It is also necessary to consider the potential reclassification of some AI systems currently defined as presenting limited risk as systems presenting considerable risk or prohibited systems.

Keywords

European Union, European Union law, artificial intelligence, legal regulation of AI, European approach, human rights

Conflict of interest	The authors declare no conflict of interest.
Financial disclosure	The study had no sponsorship.
For citation	Marchenko, A. Y., & Entin, M. L. (2022). Artificial intelligence and human rights: What is the EU's approach? *Digital Law Journal*, *3*(3), 43–57. https://doi.org/10.38044/2686-9136-2022-3-3-43-57

* Corresponding author

Submitted: 25 May 2022, accepted: 10 Aug. 2022, published: 30 Sep. 2022

СТАТЬИ

ИСКУССТВЕННЫЙ ИНТЕЛЛЕКТ И ПРАВА ЧЕЛОВЕКА: ЧТО ПРЕДЛАГАЕТ ЕВРОПЕЙСКИЙ СОЮЗ?

А.Ю. Марченко*, М.Л. Энтин

Московский государственный институт международных отношений (МГИМО-Университет) МИД России
119454, Россия, Москва, просп. Вернадского, 76

Аннотация

Интенсивное развитие технологий обозначает серьезные правовые и этические проблемы, попытки решения которых предприняты в законодательстве ЕС. Статья обращает внимание на сферу прав человека. Рассмотрены риски применения технологий искусственного интеллекта для прав человека, а также возможные варианты их преодоления. Деятельность Европейского союза здесь представляется наиболее интересной с учетом его приверженности подходу к разработке «этичного», «доверенного» ИИ, где во главе угла стоят ценности ЕС и защита прав собственных граждан. Предпосылкой для проведения настоящего исследования является передовой характер подхода ЕС к регулированию ИИ и документов, разработанных Союзом в данной области. На примере предложенного Еврокомиссией Проекта регламента по регулированию ИИ в статье анализируются положения, которые позволяют купировать или снижать данные риски, и предлагаются пути для их улучшения. Ряд предложенных правил при ближайшем рассмотрении оставляет многие чувствительные и спорные вопросы открытыми. В контексте применения технологий ИИ их решение представляется крайне необходимым, с тем чтобы интегрировать безопасный, надежный ИИ в человеческое общество. Особенно интересным данный анализ представляется за счет динамичности развития регулирования — опубликованы компромиссный текст Совета ЕС, поправки Комитета ЕС по регионам, мнения заинтересованных сторон, правозащитных организаций и другие документы, позволяющие подсветить проблемные аспекты. В статье раскрыты особенности подхода ЕС, выявлены основные достижения и пробелы, сформулированы перспективы развития регулирования ИИ.

Ключевые слова

Европейский союз, право Европейского союза, искусственный интеллект, правовое регулирование ИИ, европейский подход, права человека

Конфликт интересов	Авторы сообщают об отсутствии конфликта интересов.
Финансирование	Исследование не имело спонсорской поддержки.
Для цитирования	Марченко, А. Ю., Энтин, М. Л. (2022). Искусственный интеллект и права человека: что предлагает Европейский союз? *Цифровое право, 3*(3), 43–57. https://doi.org/10.38044/2686-9136-2022-3-3-43-57

* Автор, ответственный за переписку

Поступила: 25.05.2022, принята в печать: 10.08.2022, опубликована: 30.09.2022

Introduction

On April 21, 2021, the European Commission submitted a draft Regulation laying down harmonised rules on artificial intelligence (hereinafter, the AI Act).[1] A year later, this remains the only comprehensive document aimed at regulating almost all aspects of the creation and application of artificial intelligence technologies (hereinafter referred to as AI). The very existence of such a document represents a new stage in the regulation of AI. If adopted, it will be the first large-scale legislative act in the field of AI. Firstly, this will constitute an important example of the unification of rules at the regional level in the field of AI; secondly, due to its extraterritorial nature, it will have an impact on companies located outside the Union, third-country law, as well as the international law.

However, the adoption of this regulation is currently delayed; moreover, the original version submitted by the European Commission has already been supplemented and amended as part of the compromise text submitted by the Council of the EU under the Slovenian presidency on November 29, 2021.[2] The text of the Council of the EU introduces a number of important changes that relate to the subject matter and scope of the Act, the definition of AI systems and other definitions, prohibited uses of AI, rules for high-risk systems, as well as other key aspects that will be discussed later in the present article.

In the near future, the AI Act will undergo even more significant changes, which are to be proposed by the European Parliament (Bertuzzi & Killeen, 2022). Given the dynamic development of the AI industry, as well as disagreements over the most pressing issues, such as prohibited AI applications, certification based on self-assessment, it perhaps unsurprising that approval and adoption of the Act has been delayed. Moreover, it is also relevant to recall that, while the draft General Data Protection Regulation (GDPR) was first published in 2012, the final version was only adopted in 2016.[3] This suggests that the final adoption of the AI Act is likely to take several years.

In the context of the development and widespread use of AI technologies, it is of paramount importance that attention be paid to the sphere of human rights, which is affected directly by AI systems. In this connection, as well as considering the risks and effects of AI technologies on human rights, it is also necessary to identify possible approaches for mitigating them. Thus, in the context of its stated goals and values as one of the main defenders of the rights and freedoms of its own citizens, it is interesting to consider the EU’s approach to the regulation of AI.

This is especially relevant given the dynamic regulation environment and changes in the field of AI, which allow us to highlight controversial aspects. So, for instance, concerns expressed by many Russian and foreign experts concerning the excessively restrictive nature of the EU AI Act, upon closer examination, it turns out that many sensitive aspects have remained unresolved. For example, the classification of emotion recognition systems as posing limited risk is controversial; moreover, the list of exceptions envisaged in the ban on the use of biometric identification systems is rather broad.

Among Russian researchers, the problem of regulating the creation and use of AI technologies under the EU law is characterised by a low degree of development. Most of the relevant research

[1] 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 (April 21, 2021).

[2] Council of the European Union Presidency Compromise Text 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, (Nov. 29, 2021), https://data.consilium.europa.eu/doc/document/ST-14278-2021-INIT/en/pdf

[3] Long, W., Blythe, F., Kumar, S., & Long, W. (2022, January 18). EU Council publishes changes to Artificial Intelligence Act Proposal. *Lexology*. https://www.lexology.com/library/detail.aspx?g=717f0c32-2043-4315-ba61-9f181ace3e50

is conducted by European authors. In this connection, the large-scale works of F. Bothmer, T. Bury, N. Petit, A. Renda, R. Rodriguez, A. Siapka, and N. Smuha rate a mention. Among studies devoted to the impact of AI on human rights, the works of B. Lepri, R. Rodriguez, A. Siapka, P. Hacker should also be noted.

The present article is based on an analysis of the latest documents of the European Union in the field of AI, as well as current changes proposed by stakeholders, analytical studies, articles and reviews on the legal regulation of the creation and application of AI technologies in the EU, and the impact of AI on human rights.

The aim of the study is to identify the features, achievements and drawbacks in the EU's approach to AI regulation in the context of human rights protection and proposes approaches for improving the existing provisions.

This aim structures the following tasks:

- consider the problems and risks of human rights violations in the context of the development and implementation of AI technologies;
- analyse the latest documents of the EU, including current amendments and opinions, in the context of their ability to effectively address the above problems;
- track improvements in the original version of the AI Act in accordance with the latest changes;
- formulate the main achievements and drawbacks of the EU's approach to AI regulation and the protection of citizens' rights;
- formulate prospects for the development of legal regulation of AI technologies in the EU.

Results

After examining the main risks of human rights violations in the light of the development and implementation of AI technologies, the present study identifies problematic aspects such as lack of transparency, bias, invasiveness, and discrimination of AI systems. This provides a focus for our analysis of the AI Act and proposed amendments concerning the most controversial aspects.

The AI Act contains provisions aimed at mitigating the risk of human rights violations associated with the development of AI technologies, as well as ensuring the safe and "ethical" implementation of AI.

The AI Act envisages the prohibition of certain AI systems whose use can significantly violate human rights. This includes biometric identification systems used in public places for law enforcement purposes, social scoring systems, and a number of other types. Given that the use of such systems can be invasive and increase discrimination, their proscription seems quite reasonable.

As well as identifying high-risk systems, the AI Act encompasses a substantial number of requirements aimed at regulating the creation and implementation of these systems. For example, it is required to apply a risk management system to control risks throughout the entire life cycle of such an AI system. For this purpose, the AI Act formulates requirements for system transparency and human control.

The AI Act also contains provisions for datasets that must be up-to-date, representative, complete, error-free, and have appropriate statistical characteristics. By reducing the number of biases in datasets, this stipulation is intended to reduce the potential bias of systems and the consequent number of discriminatory decisions.

The Act also stipulates the requirement for high-risk systems to pass conformity assessments; this is designed to ensure that only those systems that meet all the necessary requirements and are safe will be allowed to enter the EU market.

The creation of a specialised supranational body and national supervisory authorities provided for in the Act is intended to facilitate coordination in the field of AI and ensure the implementation of the AI provisions. The AI Act also contains provisions on significant fines imposed in case of violation of the requirements of the Act.

A review of the proposed changes to the AI Act referring to the compromise text of the Council of the EU, the amendments of the EU Committee of the Regions, the joint position of the EDPB and EDPS, reveals several improvements compared to the original version of the AI Act:

- The changes proposed by the Council of the EU regarding the definition of AI systems provide a basis for distinguishing AI technologies from other information technologies. The compromise text does not refer to software as the only form of AI systems.
- Although the risk-based approach, which includes four levels of risk, remains unchanged, there have been clarifications regarding general-purpose AI, to which the AI Act does not apply.
- The prohibited uses of AI have been clarified. The ban on social scoring has also been extended to individuals, while the ban on the use of biometric identification systems in public places now includes the use of systems by and on behalf of law enforcement agencies, which makes it possible to extend the prohibition to include those who cooperate with law enforcement agencies.
- Annex III, which contains eight high-risk AI application areas has been updated. The following sub-items have been added: environmental protection (AI designed to control emissions and pollution) as part of Clause 2, while the AI systems used for calculating insurance premiums, underwriting, and evaluating claims are described in Clause 5 (d). Systems intended for criminal analytics are excluded from the field of law enforcement application of AI (Clause 6 (g)).
- The European Committee of the Regions point out the need for systematic notification of individuals that they interact with the system, as well as the need for such notification in relation to high-risk systems (this is not provided for in the AI Act).
- The European Data Protection Board (EDPB) and European Data Protection Supervisor (EDPS) urge that close attention be paid to emotion recognition systems, which, in their opinion, should be prohibited except in strictly defined cases (in the AI Act such systems are listed among those with limited risk).
- The EDPB and EDPS call for adaptation of the conformity assessment procedure so that preliminary assessment is always conducted by third parties in relation to high-risk systems.
- Noting the obvious improvements in comparison with the original version of the AI Act, we should point out several sensitive aspects that also require close attention:
- The AI Act and proposed amendments do not provide mechanisms for updating prohibited AI applications (Article 5) or systems with limited risk (Article 52). For high-risk AI applications, the ability to update applications is limited to specified areas. Taken together, this implies inflexibility of regulation in terms of an inability to provide a timely response to emerging threats and ensure legislative relevance to the rapidly progressing development of AI. Thus, it becomes necessary to provide for updating mechanisms and appropriate criteria.
- Manipulating and distorting people's behaviour, as well as identifying and exploiting the vulnerabilities of certain categories of citizens, would seem to comprise harmful practices that already violate human rights and thus do not require additional criteria of physical or psychological harm, as indicated in Article 5.
- Among the areas of application of high-risk AI, it is necessary to make provision for the use of AI in the healthcare sector.

- Many questions are raised by the use of AI systems for assessing the risk of an individual committing a crime or repeating it, as well as for predicting a crime or repetition thereof based on profiling or assessment of personal qualities and other characteristics. Since the AI Act considers such systems to be high-risk, it is relevant to classify such systems — or at least certain practices of their application — as prohibited.
- Since the proposed rules for high-risk AI systems are abstract by nature, they require the development of practical instructions in order to ensure their implementation in each specific case.
- It is necessary to clarify the concepts of transparency for Article 13 (high-risk AI) and Article 52 (limited-risk systems) and to include mandatory notification of a person about interaction with the AI system in Article 13.
- The compliance assessment procedure should be adapted to avoid the possibility of self-assessment, at least initially.
- Close attention should be paid to systems with limited risk (listed in Section 52: emotion recognition systems, biometric categorisation systems, deepfakes). It is important to classify some of them as prohibited (e.g., emotion recognition systems) or as high-risk, in order that they come under the appropriate regulation.
- It is important to extend to systems with limited risk the rule regarding a person's right to refuse to interact with the system in favour of a human if this is necessary to protect his or her rights.

Discussion

How AI can violate human rights. General overview

While the benefits of using AI can be significant, opening up the widest prospects for the future humanity, some AI systems and applications nevertheless involve significant risks of violating the fundamental rights of citizens. In terms of human rights, the majority of the problems associated with the use of AI and the integration of technologies into human society boil down to the risks of violating these rights. In this context, we will focus specifically on the present problems and those that may occur in near future without referring to the long-term risks of using AI, which may represent a threat to the existence of human civilisation per se. However, effective rules that allow current risks to be contained can help to prepare the ground for countering long-term threats.

Today all national and international legal documents in the field of AI emphasize the need to protect human rights. For example, in the amendments to the AI Act, the European Committee of the Regions pointed to the protection of citizens' rights as one of the goals of regulation, thus emphasising its connection with the EU Charter of Fundamental Rights.[4]

Problematic aspects may concern both the AI itself and its essence, as well as the features of its application. R. Rodriguez highlights the following problematic aspects of AI: lack of algorithmic transparency; problems associated with bias, injustice and discrimination; difficulties in challenging the decisions of AI systems; adverse impacts on the labour market; problems related to confidentiality of information and data protection (Rodriguez, 2020). These aspects are often interrelated. For instance, a lack of transparency makes it impossible to challenge the relevant decisions of a system (Edwards & Veale, 2017), while bias in datasets can lead to unfair and discriminatory decisions (Hacker, 2018).

[4] Opinion of the European Committee of the Regions on the European approach to artificial intelligence and Artificial Intelligence Act (revised opinion), 2022 O.J. (C 97) 60.

Moreover, all of these problems lead to human rights violations in one way or another. Indeed, if we consider each of the problems inherent in the AI industry such as system bias and discrimination, non-transparency of algorithms, confidentiality, data protection, and responsibility for harm caused by AI systems, it all eventually boils down to the risks of human rights violations. Such risks, which are by no means abstract, are most pronounced in particularly sensitive areas such as justice, health, public safety, employment, where the use of AI algorithms can be detrimental to human rights implying a need for protection. For instance, due to a lack of necessary transparency in AI algorithms, situations can arise where people whose rights are affected by the actions or decisions of the system do not know the reason why they were denied a particular service or why a certain decision was made in relation to them (Desai & Kroll, 2017).

Transparency of AI systems in a narrow sense means the ability to understand and explain the system's decisions. AI systems are characterised by significant complexity; moreover, a deep neural network learns independently to generate "black box effects", meaning that it is impossible to identify and explain each stage of the process in a form that is understandable to humans.[5] As a result of such opacity, the actions and decisions of AI systems often become inexplicable and untraceable, leading to the inability to prove the unfairness of the decisions made by the system, which effectively translates into the inability of citizens to protect their own rights.

Problems of injustice, bias and discrimination also become acute. Although such phenomena can be grouped together (Rodriguez, 2020) due to their significant interrelatedness, bias do not always lead to injustice or discrimination (Ferrer et al., 2021) but can remain an unnoticed deviation from the norm, which does not affect the system's decision in any way.

Typically, algorithmic bias is due to bias in datasets, which originates from the moment of data gathering and can be explained both by incorrect work with datasets and historical biases (Hacker, 2018). In recognising such data bias, algorithms can then identify additional differences to reinforce it resulting in discriminatory decisions (Siapka, 2018). Thus, the bias of algorithms is determined by existing biases in society, as well as by the diverse composition of groups working with data.

To illustrate the above-mentioned problems, we provide an example of one of the most significant EU-wide scandals involving citizens' access to social benefits in the Netherlands.[6] In 2014, with support from the Ministry of Social Affairs and Employment of the Netherlands, some cities started using the Systeem Risico Indicatie (SyRI) system, which is designed to detect fraud in the social security sector. In the process of calculating risks to predict the likelihood of fraud on the part of benefit recipients, this system collects and analyses vast amounts of data. However, people coming from the lower-income brackets of society were disproportionately evaluated, resulting in discrimination. Moreover, potential recipients of benefits did not have the opportunity to learn how the system makes decisions. In 2020, a Dutch court ruled that using the current version of SyRI is illegal due to its violation of the right to privacy in the sense described in the European Convention on Human Rights. The Court pointed out that the system was not transparent, collected too much data, and that the purposes of data collection were not clear and specific enough.[7]

[5] WIPO. Standing Committee on the Law of Patents. (2019). *Background document on patents and emerging technologies*. https://www.wipo.int/edocs/mdocs/scp/en/scp_30/scp_30_5.pdf

[6] AlgorithmWatch, (2020, April 6). *How Dutch activists got an invasive fraud detection algorithm banned*. https://algorithmwatch.org/en/syri-netherlands-algorithm/

[7] De Rechtspraak. (2020, February 13). *SyRI legislation in breach of European Convention on Human Rights*. https://www.rechtspraak.nl/Organisatie-en-contact/Organisatie/Rechtbanken/Rechtbank-Den-Haag/Nieuws/Paginas/SyRI-legislation-in-breach-of-European-Convention-on-Human-Rights.aspx

In the case of the Italian food delivery company Deliveroo,[8] a court found that the algorithm used to rank the company's couriers and determine the priority of employees when accessing convenient delivery time slots was discriminatory. In this case, the reasons why a courier did not report that he or she would not be able to go to work were not considered, meaning that a line between absenteeism and absence for valid reasons was not drawn.

Systems that track employees in the workplace are the source of various social issues due to allowing all movements and operations performed by a person, including their location, desktop screen, voice tone and other characteristics, to be recorded and analysed. In particular, various gadgets (for example, Fitbit) are used as part of so-called wellness programs for employees, transmitting data about their health to their employers (Stefano, 2018).

While the use of such vast datasets coupled with analysis tools can increase employee productivity, mitigate health and safety risks, and reduce the likelihood of accidents, such systems are programmed by humans and may not be devoid of human biases. Moreover, given the ability of AI to learn and organise itself, there is a risk that it can reprogram criteria on its own accord in order to achieve set aims, which will result in discrimination. Moreover, even if data is anonymised, the invasive collection process itself violates privacy by overstepping the boundaries of work-related processes.

In the field of labour relations, longer-term risks can also be traced. In future, the widespread use of AI systems is likely to lead to significant changes in the requirements for employees, the creation of new types of jobs, as well as inequalities in the "new" labour market.

The above-mentioned systems can effectively replace human workers currently responsible for personnel management and control. This applies not only to HR specialists, but also to employees in other fields. Over time, AI has the potential to drive humans out of many areas of activity. Thus, M. L. Entin points out that the introduction of AI in the long run does not increase human capabilities, but instead creates an alternative to them in the labour market, leaving humans with nothing to oppose (Entin & Entina, 2021).

Thus, the large-scale capabilities demonstrated by AI entail equally large-scale application risks. Numerous situations have already arisen in which human rights are violated due to the use of artificial intelligence; their number is certain to increase in the future. Therefore, attention must be paid to both current and future risks involved in the use of AI, especially in the field of human rights, as well as to develop an appropriate regulatory framework aimed at minimising such risks.

What does the EU have to offer?

Definition of AI systems. The starting point of any effective regulation is a well-defined conceptual framework. It is not an easy task to define complex, interdisciplinary, and comprehensive technologies such as artificial intelligence without narrowing or expanding the scope of regulation (Samoili et al., 2020). Moreover, the rapidly developing AI industry requires the definition to maintain its relevance even with the further development of technologies and their constant updating (Stahl et al., 2022).

Despite the rather well-developed initial wording, the definition proposed by the European Commission in the AI Act has prompted a considerable number of discussions and already undergone some changes. For instance, in the Council of the EU's compromise text, it is divided into three components. By contrast with the original text, this later version indicates the ability of such systems

[8] Allen, R., & Masters, D. (2021, January 18). *An Italian lesson for Deliveroo: Computer programmes do not always think of everything! AI-Law.* https://ai-lawhub.com/2021/01/18/an-italian-lesson-for-deliveroo-computer-programmes-do-not-always-think-of-everything/

to determine how to achieve a set of human-defined goals by training, drawing logical conclusions or modelling. According to the Council of the EU, this will permit AI technologies to be better distinguished from other information technologies. In addition to this part, the other two essentially repeat the previous version, indicating that the system receives input data (machine or "human") and generates results in the form of content, forecasts, recommendations, or decisions that affect the environment with which the system interacts. In addition, the removal from the definition of a reference to software as the only form of AI systems seems appropriate given that they may take some other form including hardware or something not yet used for such purposes.

In its amendments to the Act, the EU Committee of the Regions mentioned the impossibility of formulating a final definition of AI due to the dynamic development of the AI industry, recommending that the definition should change with the development of AI systems and applications. While committees of the European Parliament are also preparing possible amendments to the existing definition, the indications are that there will be no major changes.[9]

The EU's approach to AI regulation. The AI Act stipulates that the pan-European regulation of reliable AI will provide adequate protection for citizens and at the same time contribute to strengthening the competitiveness and production capacity of Europe in the field of AI.

Changes proposed by the Council of the EU regarding the scope of the AI Act include the exclusion of artificial intelligence systems developed for the sole purpose of conducting scientific research. Existing exclusions from the Act are AIs developed or used for national security purposes and the military.

Despite the various proposed amendments, the broad scope of the Act remains unchanged. Its provisions will also apply to companies located outside the European Union to the extent that results obtained from AI systems belonging to these companies are used in the EU. This provision remains quite controversial for many companies, who do not always know exactly where the results of their systems will be used.

The AI Act has consolidated the EU's commitment to a risk-based approach with four levels of risk: unacceptable risk, high risk, limited risk, and minimal risk. In attempting to balance the need, on the one hand, to encourage further development of innovations, and on the other hand, to protect citizens, such a risk-based approach is aimed at reducing the likelihood or extent of harm through risk assessment and regulation corresponding to the level of risk. Such a risk orientation is intended to avoid overly restrictive regulation.

However, according to representatives of several public organisations for the protection of digital rights, this approach presupposes the preliminary assignment of AI systems to various risk categories without considering that, due to its dependency on the specific context of using AI, the level of risk often cannot be fully determined in advance.[10] Moreover, such an approach is conspicuously convenient in a technical environment where companies assess their own production risks and are unlikely to be motivated to provide adequate protection of human rights.[11]

Moreover, the erroneous classification of some AI systems as low risk (for example, emotion recognition systems are assigned to this level) may lead to a lack of necessary regulation and means for

[9] Bertuzzi, L., & Killeen, M. (2022, March 4). *RT ban, internet struggles, Big Tech takes sides.* Euractiv. https://www.euractiv.com/section/digital/news/digital-brief-rt-ban-internet-struggles-big-tech-takes-sides/

[10] Statewatch. (2021, November 30). *EU: Artificial Intelligence Act must put human rights first*. https://www.statewatch.org/news/2021/november/eu-artificial-intelligence-act-must-put-human-rights-first/

[11] Hidvegi, F., Leufer, D., & Massé, E. (2021, February 17). *The EU should regulate AI on the basis of rights, not risks*. AccessNow. https://www.accessnow.org/eu-regulation-ai-risk-based-approach/

measuring such systems. In particular, in order to facilitate their access to the EU market, companies may deliberately underestimate the level of risk to avoiding complying with the rules.

Nevertheless, the EU is not likely to abandon the risk-based approach. This approach to AI systems was proposed in the White Paper on Artificial Intelligence[12] presented by the Commission in February 2020, in which only two levels of risk were proposed. At that time, the risk categories were studied more thoroughly, which eventually led to a four-tier system. It is now crucially important to clearly define the criteria for assigning systems to a certain level of risk, as well as to provide opportunities for updating for each category in order to ensure regulatory flexibility.

Prohibited uses of AI. Speaking in more detail about each level of risk, we note that within the framework of the compromise text of the Council of the EU, some changes were made to the list of prohibited AI applications (Article 5), along with updates to the areas of application of high-risk AI (Annex III). In addition, the Council of the EU has formulated a separate Article (Article 52a) for general-purpose AI capable of performing generally applicable functions such as image/speech recognition, audio/video generation, image detection, question answering, translation, and others, taking such AI to be beyond the scope of the Act.

Article 5 of the AI Act still lists prohibited uses of AI technologies (systems that distort people's behaviour and thereby harm a person or others; systems that exploit the vulnerabilities of certain groups of people; social scoring systems, remote biometric identification systems applied by law enforcement officers). Additional clarifications have been added to the first two Articles. For example, Article 1 (a) prohibits the use of systems that affect the subconscious mind and thereby distort or change people's behaviour with the aim of significantly changing a person's behaviour in a way that causes or is likely to cause physical or psychological harm to that person or another person.[13] Such formulations seem to imply that a person's behaviour can be significantly distorted without causing harm. Nevertheless, the very fact of distortion, which can take away a person's independence when it comes to decision-making, can already be considered unacceptable.

Noting that private companies such as cloud service providers are now also capable of processing vast amounts of personal data, the EDPB and EDPS insisted in their joint position issued in June 2021 on the complete ban on the use of social assessment and classification systems for individuals.[14] Such a blanket prohibition of social scoring applications demonstrates awareness of the danger such systems pose and can be interpreted as a good sign.

Significantly, the word "remote" was removed from the ban on the use of biometric identification systems in public places used for law enforcement purposes; instead, the relevant wording now indicates a "real-time". According to the definition, the collection of biometric data, comparison and identification of data in such systems have no significant delay. Moreover, this prohibition now applies to the use of these systems by or on behalf of law enforcement agencies, which makes it possible to extend it to those who cooperate with law enforcement agencies.

Exceptions to this ban remained virtually unchanged: the search for victims of crimes; the prevention of a specific threat to critical infrastructure, life, health, physical safety of individuals or a

[12] *Commission White Paper on Artificial Intelligence: A European approach to excellence and trust,* COM (2020) 65 final (February 19, 2020).

[13] Council of the European Union Presidency Compromise Text 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, (Nov. 29, 2021), https://data.consilium.europa.eu/doc/document/ST-14278-2021-INIT/en/pdf

[14] Joint Opinion of the European Data Protection Supervisor and of the European Data Protection Board on the Proposal for a Regulation of the European Parliament and of the Council laying down harmonised rules on artificial intelligence (Artificial Intelligence Act), (June 18, 2021), https://edpb.europa.eu/system/files/2021-06/edpb-edps_joint_opinion_ai_regulation_en.pdf

terrorist attack; the detection and prosecution of criminals or suspects of criminal offenses that involve a sentence of imprisonment for more than three years. Such exceptions can still be interpreted quite broadly to justify the widespread use of AI by the authorities and law enforcement agencies. The use of such systems involves processing data from a disproportionate number of subjects to identify only a few individuals, which will inevitably result in excessive invasiveness of these practices and violations of the rights of others.

The failure of the Act to address the possibility of updating the list of prohibited AI applications demonstrates the general slowness of the regulation process, reducing its ability to ensure its own relevance with the progressive development and complexity of AI-technologies. Thus, it seems appropriate to provide special mechanisms for updating prohibited AI applications (for instance, in a manner similar to the update mechanism in Annex III) so that the lists can be updated as technology evolves.

High-risk AI systems. Two groups of such systems are identified in the original version of the Act. The first category includes systems that meet two conditions: they are products or are intended to be used as components of product safety that are subject to the applicable Union legislation of Annex II (Directive 2006/42/EC on the safety of machinery and equipment, etc.), and must pass a third-party conformity assessment in order to be placed on the market or put into operation. In the compromise text, the Council of the EU only slightly changed the wording and structure of these provisions, leaving them effectively unchanged.

The second group comprises eight systems used in the areas identified in Annex III:

- biometric systems used without a person’s consent (real-time or post identification);
- critical infrastructure and environmental protection;
- education and vocational training;
- employment, employee management and access to self-employment;
- access to private and public services and benefits;
- law enforcement;
- governance migration, asylum, border control;
- administration of justice and democratic processes.

While the Act allows those specific applications of AI systems listed in the Annex can be updated by the Commission, such updates must occur within the specified eight areas.

Thus, as part of Clause 2, environmental protection (AI designed to control emissions and pollution) has been added, while Clause 5 (d) now specifies AI systems used in insurance for calculating insurance premiums, underwriting, and evaluating claims.

As for biometric identification, in general this practice is associated with a substantial risk of invasion of people’s privacy and violation of anonymity. While the phrase “without a person’s agreement” implies that, in order to use such systems, the person must be informed, the problem of properly informing individuals about such processing has not yet been solved.

Thus, the EDPB and EDPS call for a ban on any use of AI for automatic recognition of human features in public places (faces, gaits, fingerprints, DNA, etc.), as well as systems for biometric categorisation of people. In their joint position, it is also noted that, when it comes to political gatherings and protests, subsequent identification can adversely affect rights and freedoms such as freedom of assembly and association, as well as the fundamental principles of democracy.[15]

[15] Joint Opinion of the European Data Protection Supervisor and of the European Data Protection Board on the Proposal for a Regulation of the European Parliament and of the Council laying down harmonised rules on artificial intelligence (Artificial Intelligence Act), (June 18, 2021), https://edpb.europa.eu/system/files/2021-06/edpb-edps_joint_opinion_ai_regulation_en.pdf

It is quite problematic to use AI for individual assessments of the risk of committing a crime or the repetition thereof by an individual, as well as for predicting a crime based on profiling or assessment of personal qualities, past behaviour, etc. The use of algorithms to predict future human behaviour in such sensitive areas will inevitably put lower-income brackets of society at risk, leading to discrimination, which in some cases may destroy lives. Therefore, such systems or individual practices of their application, should be classified as prohibited.

The main part of the Act is aimed specifically at regulating high-risk AI. In relation to such systems, suppliers are supposed to provide a risk management system throughout the entire system life cycle, take measures to eliminate or mitigate risks, and conduct post-market monitoring. The Act requires systems to be designed in such a way as to ensure record-keeping during their operation (Article 12), provide transparency of the AI system, ensure availability of information (Article 13). Human control should be conducted throughout the entire life cycle of the system, including the ability to interfere with the system at any time and, if necessary, stop or adjust it (Article 14).

Despite the obvious rationality of the proposed rules, they are rather abstract and require additional practical documents and instructions to ensure compliance in each specific case. The conformity assessment procedure can be based on internal control (self-assessment; Annex VI) or conducted with the participation of a third party (Annex VII). While systems that are subject to EU legislation (Annex II) are also subject to a conformity assessment procedure in accordance with these acts, an AI compliance assessment must be part of this assessment.

With the exception of biometric identification, all of the systems in Annex III are subject to an assessment procedure conducted without the participation of an authorised body. This raises a number of questions. In general, the self-assessment procedure remains controversial due to its potentially insufficiency in terms of protecting human rights.

B. Benifei, the representative of the European Parliament's Internal Market and Consumer Protection Committee (IMCO), referred to the potential failures of such a procedure: "We don't want to detect biases in systems after they've already destroyed families and lives, as has happened in some countries."[16] In their joint position, the EDPB and EDPS also point to the need to adapt the conformity assessment procedure for high-risk systems so that the assessment is always carried out by third parties.

Transparency. While the transparency of AI systems can be difficult to achieve in some cases, it is precisely transparency that can increase citizens' confidence in such systems, as well as provide the ability to control the AI. When this requirement clashes with a trade secret regime, it is important to ensure that the systems are as transparent as possible, at least to the competent supervisory authorities, otherwise we risk finding ourselves in a situation of unexplained actions and decisions that can significantly affect human lives.

Despite the Act containing two Articles on the transparency of AI systems, the uniform wording implies different requirements. Thus, Article 13 (transparency of high-risk systems) deals with the possibility of the interpretation of output data of the system for further use, while Article 52 (transparency of systems with limited risk) refers to the need to inform an individual subject about his or her interaction with the system.[17]

While the Act requires people to be informed when they interact with the AI systems listed in Article 52, it does not contain a similar requirement for high-risk systems that pose an even greater

[16] Bertuzzi, & Killeen, 2022.

[17] Kiseleva, A. (2021, July 29). *Making AI's Transparency Transparent: AI'S: notes on the EU Proposal for the AI Act (2021).* European Law Blog. https://europeanlawblog.eu/2021/07/29/making-ais-transparency-transparent-notes-on-the-eu-proposal-for-the-ai-act/

threat. It remains unclear what information should be specified in such a notification, for example, whether or not information should be included regarding the goals, the logic of the intended actions, or the right to request explanations.

The exception to Article 52 comprises systems used for the detection, prevention, investigation or prosecution of criminal offences. Here again, one cannot fail to note the excessive breadth of the exception. Here it is advisable to distinguish the area of detection and prevention of crimes as requiring greater guarantees for the protection of citizens that take into account the presumption of innocence. It is also possible to cite opinions about the need to completely cancel these exceptions, since the use of such manipulative AI systems without ensuring the required transparency presents a serious threat to fundamental rights.[18]

It should be noted that Article 52 does not contain provisions regarding the possibility of a person to refuse to interact with the system in favour of interacting with a person if this is necessary to protect fundamental rights, as is set out in the Ethics Guidelines for Trustworthy AI prepared by HLEG AI. However, according to the amendments to the Act proposed by the EU Committee on Regions, the scope of opportunities and legal status of individuals interacting with AI systems should not be limited to this interaction.

The EU Committee of the Regions also pointed out the need to systematically notify individuals that they are interacting with the system. The AI Act indicates that individuals should be notified only in cases where this is not obvious from the circumstances and context of use. The Committee also noted: "Natural persons should always be duly informed whenever they encounter AI systems, and this should not be subject to the interpretation of the given situation."[19]

Thus, regardless of whether or not this is obvious from the circumstances of using AI, individuals should always be informed about interactions with AI systems; moreover, this requirement should be extended to high-risk systems that are used in particularly sensitive areas.

Limited-risk systems. Article 52 of the AI Act is aimed at regulating systems with limited risk. These systems, in the opinion of the Commission, do not pose such a significant risk as to be classified as high-risk systems. However, we will try to find out whether these systems are indeed so harmless that they do not require all the complex procedures that are applicable to high-risk AI.

Therefore, the Article specifies the following systems: emotion recognition systems, biometric categorisation systems, deepfakes (systems that generate or manipulate images, audio or video content, real people, objects, places, or other objects or events).

Emotion recognition systems collect and process data and information about a person's mental processes. Examples of such systems can be found in China, where human rights organisations report that Uighurs are undergoing experiments aimed at determining their emotional state.[20] Similar systems can also be used in marketing and social networks aimed at influencing and manipulating people's behaviour in order to mislead them. However, the fact that such systems are not listed as prohibited or high-risk seems rather questionable given the level of risk and questionable use of some of them.

[18] Statewatch. (2021, November 30). *EU: Artificial Intelligence Act must put human rights first*. https://www.statewatch.org/news/2021/november/eu-artificial-intelligence-act-must-put-human-rights-first/

[19] Opinion of the European Committee of the Regions on the European approach to artificial intelligence and Artificial Intelligence Act (revised opinion), 2022 O.J. (C 97) 60.

[20] Malgieri, G., & Ienca, M. (2021, July 7). *The EU regulates AI but forgets to protect our mind.* European Law Blog. https://europeanlawblog.eu/2021/07/07/the-eu-regulates-ai-but-forgets-to-protect-our-mind/

Biometric categorisation systems are those aimed at dividing people into categories depending on their ethnicity, gender, political views, sexual orientation etc., based on their biometric data. Deepfakes are fake videos and images created with the intention to discredit or mislead.

According to the Act, it is only possible to provide protection by notifying a person about his or her interaction with the system without providing the opportunity to refuse such interaction. At the same time, the provision concerning the possibility of refusing to interact with the system in favour of interacting with a person is necessary in the context of protecting fundamental rights and ensuring non-discrimination.

Thus, the use of such systems remains insufficiently regulated, except in rare cases when their use causes mental or physical harm or exploits the vulnerabilities of certain groups of people, in which case it will be considered as prohibited.

The need to pay close attention to emotion recognition systems is noted by EDPB and EDPS in their joint statement. Representatives have indicated that the use of emotion recognition systems is highly undesirable and should be prohibited, except for clearly defined applications such as medical applications, where it is necessary to the recognise the emotional state of a patient.[21]

Conclusion

Comprising an all-embracing and ubiquitous technology, artificial intelligence has raised a number of issues that need to be addressed immediately. In order to prepare the ground for overcoming long-term threats, it is of immense importance to quickly and efficiently work out the risks and threats that AI carries.

As one of the leaders in AI regulation, The EU is definitely on the right track, working out comprehensive standards, consulting with all stakeholders in order that the version of the AI Act that is adopted into law takes all relevant factors into account. The post-covid disunity and heterogeneity of the EU member states may be gradually being replaced with a sense of cohesion (Entin & Entina, 2021). Such cohesion of the EU member states is necessary when it comes to regulating the creation and use of AI systems to ensure the protection of its citizens.

The AI Act proposed by the European Commission in April 2021 sets out a number of regulations designed to minimise the threats and risks associated with the rapid development of AI technologies. Once adopted, its provisions will obviously be reflected in both international law and the third-country law; moreover, the Act itself will significantly affect the global artificial intelligence market.

References

1. Desai, D. R., & Kroll, J. A. (2017). Trust but verify: A guide to algorithms and the law. *Harvard Journal of Law & Technology, 31*, 1–64.
2. Edwards, L., & Veale, M. (2017). Slave to the algorithm? Why a ‘Right to an Explanation’ is probably not the remedy you are looking for. *Duke Law and Technology Review*, *16*(1), 18–84. https://scholarship.law.duke.edu/dltr/vol16/iss1/2

[21] EDPB Press Release Statement (2021, June 2021). *EDPB & EDPS call for ban on use of AI for automated recognition of human features in publicly accessible spaces, and some other uses of AI that can lead to unfair discrimination.* https://edpb.europa.eu/news/news/2021/edpb-edps-call-ban-use-ai-automated-recognition-human-features-publicly-accessible_en

3. Entin, M. L., & Entina, E. G. (2021). *V poiskah partnerskih otnoshenij — X: Rossiya i Evropejskij soyuz v 2020 — pervoj polovine 2021 godov* [Looking for partnership — X: Russia and the European Union in 2020 — the first half of 2021]. Zebra E.
4. Ferrer, X., Nuenen, T., Such, J. M., Coté, M., & Criado, N. (2021). Bias and discrimination in AI: A cross-disciplinary perspective. IEEE Technology and Society Magazine, *40*(2), 72–80. https://doi.org/10.1109/MTS.2021.3056293
5. Hacker, P. (2018). Teaching fairness to artificial intelligence: Existing and novel strategies against algorithmic discrimination under EU Law. *Common Market Law Review, 55*(4), 1143–1185. https://doi.org/10.54648/cola2018095
6. Rodrigues, R. (2020). Legal and human rights issues of AI: Gaps, challenges and vulnerabilities. *Journal of Responsible Technology*, *4*, Article 100005. https://doi.org/10.1016/j.jrt.2020.100005
7. Samoili, S., López Cobo, M., Gómez, E., De Prato, G., Martínez-Plumed, F., & Delipetrev, B. (2020). *AI Watch defining artificial intelligence. Towards an operational definition and taxonomy of artificial intelligence*. Joint Research Centre. https://publications.jrc.ec.europa.eu/repository/handle/JRC118163
8. Siapka, A. (2018). *The ethical and legal challenges of artificial intelligence: The EU response to biased and discriminatory AI*. SSRN. http://dx.doi.org/10.2139/ssrn.3408773
9. Stahl, B., Rodrigues, R., Santiago, N., & Macnish, K. (2022). A European Agency for Artificial Intelligence: Protecting fundamental rights and ethical values. *Computer Law & Security Review, 45*, Article 105661. https://doi.org/10.1016/j.clsr.2022.105661
10. Stefano, V. (2018) *"Negotiating the algorithm": Automation, artificial intelligence and labour protection. Emloyment*. Working Paper No. 246. International Labour Office, Geneva. https://www.ilo.org/wcmsp5/groups/public/---ed_emp/---emp_policy/documents/publication/wcms_634157.pdf

Information about the authors:

Anna Y. Marchenko* — Ph.D. Student, Department of European Law, MGIMO-University, Moscow, Russia.
anna.yur.marchenko@gmail.com
ORCID: https://orcid.org/0000-0003-1601-7432

Mark L. Entin — Dr. Sci. in Law, Professor, Head of European Law Department, MGIMO-University, Moscow, Russia.
entinmark@gmail.com
ORCID: https://orcid.org/0000-0001-9562-8340

Сведения об авторах:

Марченко А. Ю.* — аспирант кафедры европейского права Московского государственного института международных отношений (МГИМО-Университет) МИД России, Москва, Россия.
anna.yur.marchenko@gmail.com
ORCID: https://orcid.org/0000-0003-1601-7432

Энтин М. Л. — доктор юридических наук, профессор, заведующий кафедрой европейского права Московского государственного института международных отношений (МГИМО-Университет) МИД России, Москва, Россия.
entinmark@gmail.com
ORCID: https://orcid.org/0000-0001-9562-8340