THE PROTECTION OF HUMAN RIGHTS UNDER THE ARTIFICIAL INTELLIGENCE ACT

Pham Thi Minh Trang
Lecturer, Commercial Department, Ho Chi Minh City University of Law (Vietnam)
Doctorate Student, Doctoral School of Law, University of Pécs (Hungary)
Corresponding address: ptmtrang@hcmulaw.edu.vn
ORCID: 0009-0002-2291-2724

DOI: 10.47272/KIKPhD.2025.1.2

ABSTRACT

Artificial intelligence (AI) has developed rapidly and is having a profound impact on society as a whole. AI-driven technologies are now present in sectors such as healthcare, agriculture, food safety, education, media, sports, and culture, where they have the potential to optimize human time and enhance work efficiency. However, the social prospects of AI are both alluring and alarming; its promises and perils are difficult to disentangle. The risks to users have become an increasing concern as AI technologies are embedded in everyday products and services. Furthermore, AI can influence human behavior in new and unexpected ways, potentially undermining human dignity. To ensure better conditions for the application and use of AI in the development of social and economic sectors – while placing human rights at the center – the European Union enacted the AI Act, which entered into force in August 2024. This is the world's first comprehensive AI legislation, establishing a legal framework for both users and developers of AI systems in Europe. It aims to create a safe, transparent, and trustworthy environment for the deployment of AI technologies.

This paper will examine AI systems and explore the current challenges related to human rights in the context of rapid AI advancement. In addition, it will analyze the provisions of the AI Act to shed light on how it addresses the protection of human rights.

KEYWORDS

Human rights, fundamental rights, artificial intelligence, AI Act, AI systems.

ARTICLE HISTORY

SUBMITTED 28 Dec 2024 | REVISED 18 Feb 2025| ACCEPTED 28 Feb 2025

I. Introduction

The development of technology and the integration of AI into everyday life are accelerating rapidly. AI-driven technologies are increasingly permeating individuals' daily routines, from smart home appliances to social media applications.[1] Public authorities are also increasingly using AI to assess individuals' personalities or competencies.[2] AI enables technical systems to perceive their environment, process information, solve problems, and act autonomously to achieve specific goals.[3]

A major milestone in AI development was the launch of a groundbreaking chatbot system by OpenAI, which marked a significant turning point in the development of AI. This system possesses an extensive knowledge base and can engage in conversations across a wide range of fields, including technology, science, business, education, and more. It can understand context, generate human-like dialogue, and improve continuously through user interaction.

In general, AI represents an innovative digital system capable of self-learning, developing its own search and learning algorithms, constructing artificial neural networks, and even writing its own code. Most importantly, it possesses decision-making capabilities.[4] As such, AI increasingly affects many aspects of life and poses potential threats to a wide array of human rights, including the right to non-discrimination, freedom of expression, human dignity, personal data protection, and privacy.[5] Notably, AI is not an isolated issue,[6] it is inextricably linked to the processing of personal data and privacy concerns. During the use of AI systems, users often provide significant amounts of personal information, which may be accessed and used without their explicit consent. This raises substantial risks and may harm both public interests and individual rights.

Moreover, AI - 'decision-makers' can directly affect the health and life of one or more humans. For instance, AI plays a crucial role in the development of autonomous (driverless) vehicles, which operate without human intervention. In

[1] For example, the online sales system uses AI to analyze user habits, helping to optimize advertising content for each customer. Therefore, similar products that you have searched for will continue to appear in ads on your phone or laptop.

[2] Council of Europe Commissioner for Human Rights, *Unboxing Artificial Intelligence: 10 steps to protect Human Rights* (2019) https://rm.coe.int/unboxing-artificial-intelligence-10-steps-to-protect-human-rights-reco/1680946e64 accessed 1 March 2025.

[3] European Parliament, 'What is artificial intelligence and how is it used?' (2020) https://www.europarl.europa.eu/topics/en/article/20200827STO85804/what-is-artificial-intelligence-and-how-is-it-used accessed 1 March 2025.

[4] Ingrid Lleana Nicolau, 'Human Rights and Artificial Intelligence' (2019) 12 *Journal of Law and Administrative Sciences* 64.

[5] European Parliament, *Artificial Intelligence Act: Briefing EU Legislation in Progress* (2024) https://www.europarl.europa.eu/RegData/etudes/BRIE/2021/698792/EPRS_BRI(2021)698792_EN.pdf accessed 1 March 2025.

[6] Chris Lewis, 'The Need for Legal Framework to Regulate the Use of Artificial Intelligence' (2022) 47(2) *University of Dayton Law Review* 287.

traffic situations—especially when a crash is imminent—human lives depend on split-second decisions made by software-driven systems. This highlights the urgent need for regulatory frameworks and ethical guidelines to define the permissible scope of AI decision-making. Without sufficient transparency, AI systems may produce biased or erroneous outcomes.[7] Therefore, human oversight is essential to prevent harmful consequences.

AI should serve as a tool for human benefit, contributing positively to society with the ultimate aim of enhancing human well-being.[8] Addressing the challenges of opacity, complexity, bias, unpredictability, and partial autonomy in certain AI systems is essential to ensure compatibility with fundamental rights and facilitate the enforcement of legal standards.[9] Recognizing both the benefits and risks of AI, EU policymakers have adopted a "human-centric" approach to AI,[10] aiming to maximize the benefits of new technologies while mitigating their associated risks. At the same time, strengthening the legal framework is an immediate priority, as it is essential for guiding AI development and ensuring a safe and rights-respecting environment.

In 2018, the European Commission published its AI strategy and established the 'High-level Expert on Artificial Intelligence' (AI HLEG) to support its implementation.[11] The foundation of AI regulations began with the Ethics Guidelines for Trustworthy AI in 2018, which identified three essential components of trustworthy AI: legality, ethical alignment, and technical robustness. In 2019, the Commission issued these guidelines in a non-binding form as a soft law instrument to provide operational guidance. Subsequently, the EU moved toward adopting harmonized rules for the development, market placement, and use of AI systems. In February 2020, the Commission launched its White Paper on AI, initiating a public consultation on the future regulatory framework.[12]

[7] Carol M Bast, 'Artificial Intelligence and Ethics' (2024) 50(2) *Rutgers Computer and Technology Law Journal* 285.

[8] 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.

[9] Council of the European Union, *Presidency Conclusions – The Charter of Fundamental Rights in the Context of Artificial Intelligence and Digital Change*, 11481/20 (2020) https://www.consilium.europa.eu/media/46496/st11481-en20.pdf accessed 1 March 2025.

[10] European Parliamentary Research Service, *Artificial Intelligence Act: EU Legislation in Progress – Briefing* (2024) https://www.europarl.europa.eu/RegData/etudes/BRIE/2021/698792/EPRS_BRI(2021)698792_EN.pdf accessed 1 March 2025.

[11] Martin Ebers, 'Standardizing AI – The case of the European Commission's Proposal for an "Artificial Intelligence Act"' in Larry A Dimatteo, Cristina Poncibò and Michel Cannarsa (eds), *The Cambridge Handbook of Artificial Intelligence: Global Perspectives on Law and Ethics* (Cambridge University Press 2022) 325.

[12] European Commission, *White Paper on Artificial Intelligence: A European Approach to Excellence and Trust* (2020) https://commission.europa.eu/publications/white-paper-artificial-intelligence-european-approach-excellence-and-trust_en accessed 1 March 2025

This process culminated in the enactment of the AI Act (Regulation 2024/1689), which entered into force in 2024.[13] It is the world's first binding, horizontal regulation on AI, establishing a common legal framework for the development and deployment of AI systems within the EU.[14] he regulation underscores the EU's dual focus on innovation and economic growth, as well as on ethical and societal implications. Compliance with European ethical principles, legal standards, and social values is framed as essential to building 'an ecosystem of trust'.[15]

This paper employs legal analysis, commentary, and comparative methods to clarify the following issues. First, it provides a historical overview of the definition of artificial intelligence and its development over time, alongside a discussion of relevant legal provisions. Second, it categorizes AI systems based on their risk levels. Third, it analyzes how AI systems may affect human rights and evaluates the AI Act's mechanisms for safeguarding these rights.

II. Overview of the Key Characteristics of Artificial Intelligence

It is important to have a clear understanding of AI that captures aspects relevant to societal intervention while being aware of AI's current and future technical capabilities.[16] There have been many different definitions of AI over the years, but there is no standard definition of what AI actually involves[17]. In 1955, a Dartmouth mathematics professor named John McCarthy declared that "every aspect of learning or any other feature of intelligence can in principle be so precisely described that a machine can be made to simulate it".[18] Then, in 1956, the term 'artificial intelligence' was officially coined.[19] According to Bellman (1978), AI is defined as "the automation of activities that we associate with human thinking, such as decision-making, problem-solving, and learning."[20] Kurzweil (1990) further

[13] 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) [2024] OJ L 2024/1689, (Artificial Intelligence Act).

[14] European Parliamentary Research Service (n 10).

[15] Ebers (n 11) 325.

[16] Pascal D König, Tobias D Kraff, Wollfganf Schulz and Katharina A Zweig, 'Essence of AI – What is AI?' in Larry A Dimatteo, Cristina Poncibò and Michel Cannarsa (eds), *The Cambridge Handbook of Artificial Intelligence: Global Perspectives on Law and Ethics* (Cambridge University Press 2022) 18. https://doi.org/10.1017/9781009072168.005

[17] JRC Technical Reports, *AI Watch – Defining Artificial Intelligence* (European Commission, 2019) https://publications.jrc.ec.europa.eu/repository/bitstream/JRC118163/jrc118163_ai_watch._defining_artificial_intelligence_1.pdf accessed 1 March 2025.

[18] Lewis (n 6) 288.

[19] Rockwell Anyoha, 'The History of Artificial Intelligence' (2006) https://sitn.hms.harvard.edu/flash/2017/history-artificial-intelligence/ accessed 1 March 2025.

[20] Anne Bowser, Michael Sloan, Pietro Michelucci and Eleonore Pauwels, *Artificial Intelligence: A Policy-Oriented Introduction* (Wilson Center, 2017)

described AI as "the art of creating machines that perform functions which, when performed by people, require intelligence,"[21] while Rich and Knight (1991) framed it as "the study of how to make computers do things which, at the moment, people are better at."[22] In short, in simple terms, AI is the ability of a machine to display human-like capabilities such as reasoning, learning, planning and creativity.[23] It can process input data into recognizable patterns and/or then use those patterns to formulate decisions.[24]

From a legal perspective, as Article 3 of the AI Act, *'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. This term refers to general-purpose AI (GPAI) models 'that are trained with a large amount of data using self-supervision at scale', that display 'significant generality' and are 'capable to competently perform a wide range of distinct tasks' and 'can be integrated into a variety of downstream systems or applications'.[25] In other words, AI systems run by complicated algorithms and wide-ranging data collected during the training process. Besides, AI exhibits a level of proficiency comparable to a human expert in most fields. In particular, its capabilities in various areas, such as communication, self-learning, adaptation, decision-making, and generating new outputs, have made AI increasingly advanced.

It can be seen that the legal definition of 'AI system' is quite broad, which significantly widens the scope of the AI Act. The regulations apply directly to public and private actors both inside and outside the EU as long as the AI system is applied to the EU market or its use affects people located in the EU.[26] Article 2(1) lists the subjects within the scope of the AI Act, such as providers and deployers of AI systems, importers and distributors of AI systems, product manufacturers, and affected persons in the Union, to name a few. However, the AI Act's scope has several exemptions for AI systems, such as those used for military or defense

https://www.wilsoncenter.org/sites/default/files/media/documents/publication/wilson_center_policy_brief_artificial_intelligence.pdf accessed 1 March 2025.

[21] Frank Emmert-Streib, Olli Yli-Harja and Matthias Dehmer, 'Artificial Intelligence: A Clarification of Misconceptions, Myths and Desired Status' (2020) 3 *Front Artif Intell* 524339 https://doi.org/10.3389/frai.2020.524339

[22] Department of Aeronautical Engineering, *Artificial Intelligence* (Malla Reddy College of Engineering & Technology) https://mrcet.com/downloads/digital_notes/AE/III/Artificial%20Intelligence.pdf accessed 1 March 2025.

[23] European Parliament (no 2)

[24] *Further reference:* Lewis (n 6) 286-311.

[25] European Parliamentary Research Service (n 10)

[26] Ebers (n 11) 333.

purposes, and limited exemptions for free and open-source systems.[27] For example, in Recital 24 of the AI Act, AI systems placed on the market, put into service, or used—with or without modification—for military, defense, or national security purposes should be excluded from the scope of this Regulation, regardless of whether the entity carrying out those activities is public or private. This exclusion is justified by the fact that national security remains the sole responsibility of Member States, as well as by the specific nature and operational needs of national security activities, along with the national rules applicable to those activities. Or, in Recital 25, the AI act will not apply to AI systems and models, including their output, which are specifically developed and put into service for the sole purpose of scientific research and development.[28]

Furthermore, this definition has also been the subject of controversy due to its broadness.[29] The EU Commission's original legislative proposal referred to 'software' for AI systems developed using certain specifically defined 'techniques and concepts' (such as deep learning, inference and deduction machines, and statistical approaches). However, there is a problem with this approach: it was too broad in its use of 'software' and too narrow in specifying particular 'techniques and concepts', making it overall inaccurate.[30] In Article 3 of the AI Act, the term *'machine-based'* is used to refer to the fact that AI systems are operated by machines. However, the AI Act does not provide a specific definition of what is meant by 'machine' in this context.

It is evident that regulating the definition of AI systems is significantly challenging. As AI is notoriously difficult to define.[31] There are several reasons why it is difficult to arrive at a unified definition. Many subfields are researching AI, each with various approaches and methods. Moreover, the disciplinary heterogeneity of AI research is further increased by its ties to other fields, such as neuroscience, biology, and cognitive science. Besides, many definitions of AI are not widely recognized among researchers. They are not wrong, but they are not useful.[32] In the meantime, AI is growing and becoming more complex. Therefore, the demand for

[27] KPMG, *Decoding the EU AI Act: Understanding the AI Act's Impact and How You Can Respond* (2023) https://assets.kpmg.com/content/dam/kpmgsites/xx/pdf/2024/02/decoding-the-eu-ai-act.pdf accessed 1 March 2025.

[28] *Further reference:* European Parliamentary Research Service (n 10)

[29] Rubén Cano, 'A Proposal for (AI) Change? A Succinct Overview of the Proposal for Regulation Laying Down Harmonised Rules on Artificial Intelligence' (2021) https://iplens.org/category/artificial-intelligence/ accessed 1 March 2025.

[30] Lukas Feiler, Alexander Hofmann and Beat König, *'AI Act: Regulation with Little Accuracy' (2024) IT Law Decoded Blog https://www.derstandard.at/story/3000000203623/ai-act-regulierung-mit-wenig-treffsicherheit* accessed 1 March 2025.

[31] König et al (no 6) 23.

[32] Pei Wang, 'On Defining Artificial Intelligence' (2019) 10(2) *Journal of Artificial General Intelligence* 5 https://doi.org/10.2478/jagi-2019-0002

exactness can only be relatively satisfied, as there is no way to completely remove ambiguity in a definition.[33]

The AI Act follows a risk-based approach, categorizing AI systems into four risk levels: Unacceptable risk (Prohibited AI practices), High-risk AI system, Limited risk (Transparency risk), and Minimal risk. Each type is categorized according to different standards. Based on the level of risk, an AI system must comply with various requirements to ensure the development of trustworthy AI and to minimize risks. This approach is illustrated by Recital 26, "a clearly defined risk-based approach should be followed" to create a proportionate and effective set of binding rules for AI systems. The categories are a horizontal approach, meaning they are not sector-specific but pertain to the broader, general use of AI.

(i). Unacceptable risk: The AI Act bans applications and systems that create an unacceptable risk. The prohibited AI practices contravene Union values, and their harm is unacceptable due to their threat to the safety, livelihood and rights of individuals[34]. In other words, Article 5(1) lists the use of AI, which poses a high potential risk of violating fundamental rights and social values. In the AI Act, the scope of prohibited practices is broader than before and is regulated in Article 5(1). Some typical examples include AI systems using subliminal techniques that are out of a person's consciousness, or purposefully manipulative or deceptive techniques, which lead to significant harm to a person or a group of persons. Because the adverse effects impair human judgment or decision-making ability. Another concern is that AI systems exploit any of the vulnerabilities of a person or groups based on their age, disability, or a specific social or economic situation. The regulation also prohibits the use of AI to exploit personal data illegally or infer sensitive attributes such as race, political opinions, through the use of biometric categorisation systems.

(ii). *High-risk:* cover high-risk applications and systems because they can potentially create an adverse impact on people's health, safety, environment or fundamental rights.[35] High–risk AI systems are permitted on the European market but are subject to compliance with mandatory requirements and conformity assessment[36] before they can be launched on the market. In other words, before high-risk AI systems can be put into service or used on the Union market, they have to comply with mandatory requirements regulated by the AI Act. Those requirements aim to ensure that the use of high-risk AI systems do not give rise to any unacceptable risks. As a result, public interests and human dignity in the Union are protected and safeguarded by Union law. High-risk AI systems are regulated in Article 6.2 of the AI Act and Annex III. Some sectors with high-risk AI systems include the following: evaluation of eligibility for credit, health or life insurance, or

[33] ibid 5.
[34] Ebers (n 11) 334.
[35] European Parliament (n 10).
[36] 'Conformity assessment' means the process of demonstrating whether the requirements set out in Chapter III, Section 2 of AI Act relating to a high-risk AI system have been fulfilled;

public benefits; analysis of job applications or candidate evaluations; and product safety components.

(iii) Limited risk: Users must be aware that they are communicating with or dealing with AI systems. In other words, AI systems with limited risks designed to interact directly with natural persons must inform the individuals concerned that they are interacting with an AI system. Moreover, AI systems with limited risks can generate synthetic audio, image, video, or text content, and such outputs should be labeled clearly as AI-generated or manipulated. This concerns AI systems that interact with humans (chatbots), emotion recognition, and biometric categorization systems and systems that generate or manipulate content[37] (deep fakes[38]). These systems shall adhere to transparency and information requirements.[39] Obligations for limited-risk systems focus on outputs and users.

(iv). *Minimal risks:* those AI systems that do not fall into the three categories above. Therefore, there are no requirements to meet any obligations.[40] For example, AI-enabled video games or spam filters. Nevertheless, the EU strongly encourages the development of codes of conduct to foster the wider adoption of reliable AI.

III. Artificial Intelligence and the Protection of Human Rights: Challenges and Solutions

In the simplest terms, human rights can be defined as the collective and individual rights which are to be enjoyed by every human being by the virtue of their birth.[41] As outlined by the United Nations Office of the High Commissioner for Human Rights, the human rights that may be adversely affected by generative AI, listed in the order they appear in the Universal Declaration of Human Rights (UDHR), these include the right to freedom from physical and psychological harm; the right to equality before the law and protection against discrimination; the right to privacy; the right to own property; the right to freedom of thought, conscience, religion, and opinion; the right to freedom of expression and access to information; the right to participate in public affairs; the right to work and earn a living; the rights of the

[37] Ebers (n 11) 334.

[38] AI Act, art 3(60). *'Deep fake'* means AI-generated or manipulated image, audio or video content that resembles existing persons, objects, places, entities or events and would falsely appear to a person to be authentic or truthful

[39] Balázs Hohmann, 'The Interpretation of Transparency from the Legal Point of View' in Tamás Haffner (ed), *4th Youth in Europe Conference – Proceedings* (Sopianae Cultural Association 2018) 155–163
Balázs Hohmann, 'Interpretation of the Concept of Transparency in the Strategic and Legislative Documents of Major Intergovernmental Organizations' (2021) 2(1) *PhD Studies in Administrative and Infocommunications Law* 50-54. https://doi.org/10.47272/KIKPhD.2021.1.4

[40] Forvis Mazars, 'EU AI Act: Different Risk Levels of AI Systems' https://www.forvismazars.com/ie/en/insights/news-opinions/eu-ai-act-different-risk-levels-of-ai-systems accessed 1 March 2025.

[41] *Universal Declaration of Human Rights* (adopted 10 December 1948 UNGA Res 217 A(III)).

child; and the right to culture, art, and science.[42] It's clear that the scope of AI may impact human rights broadly. In this paper, *inter alia,* the focus will be on a select few human rights affected by the AI Act, rather than all the affected rights.

As mentioned before, with a horizontal approach, the AI Act does not regulate sector-specific areas but categorizes AI systems by risk. The management of AI systems becomes more flexible and effective through this method. For each type of AI system, this Act lays down the fundamental requirements and necessary safeguarding mechanisms. Moreover, it emphasizes a high level of traceability and transparency obligations. In the first recitals of this Act, it is stated that the regulation aims to boost innovation and the development of AI systems that are trustworthy, human-centric, and respectful of human rights, thereby achieving the goal of balancing AI development and innovation with the effective management of emerging risks. Additionally, it seeks not only to prevent harmful or abusive acts but also to protect human dignity, freedom, equality, democracy, and fundamental rights.

First, AI systems affect privacy and personal data. The Universal Declaration of Human Rights declared that *"No one shall be subjected to arbitrary interference with his privacy, family, home or correspondence, nor to attacks upon his honour and reputation. Everyone has the right to the protection of the law against such interference or attacks."* (Article 12) However, generative AI raises several concerns related to the right to privacy.

The longest and most sustained human rights debate on automated data processing and algorithms relates to the right to privacy.[43] This issue derives from a massive data source necessary for the operation of AI systems. In other words, the input data is the essential factor for their functionality and performance. Through a series of algorithms and data, AI systems can solve prompts, create new outputs, make conversations, or make prediction. These algorithms are mathematical, logical commands with step-by-step instructions on how to process the input data. Thus, AI systems must collect extensive data from users, especially personal data to learn and perform tasks. This data can include personal information such as names, addresses, financial information, and sensitive information such as medical records and social security numbers.[44] During the process of operation or self-assessment, the AI system poses a real risk of violating human rights. Main concerns include how the collected data is being used, who has access to it, and whether this data is being abused or misused. Depending on its enormous amount of data, AI systems may create harmful, false and convincing content that may be used to directly attack

[42] United Nations Human Rights Office of the High Commissioner, *Taxonomy of Human Rights Risks Connected to Generative AI* https://www.ohchr.org/sites/default/files/documents/issues/business/b-tech/taxonomy-GenAI-Human-Rights-Harms.pdf accessed 1 March 2025.

[43] Arthur J Sills, 'Automated Data Processing and the Issue of Privacy' (1970) 1 *Seton Hall L Rev* 7.

[44] The Economic Times, 'AI and Privacy: The Privacy Concerns Surrounding AI, Its Potential Impact on Personal Data' (2023) https://economictimes.indiatimes.com/news/how-to/ai-and-privacy-the-privacy-concerns-surrounding-ai-its-potential-impact-on-personal-data/articleshow/99738234.cms?from=mdr accessed 1 March 2025.

an individual's privacy, honour or reputation.[45] For example, these data and information can be exploited for anti-social purposes or fraudulent activities. A fraudster might use personal details from social media profiles or a resume, then leverage AI applications to impersonate individuals and carry out extortion via traditional phone calls.

The protection of privacy and personal data is adjusted by both the AI Act and the EU General Data Protection Regulation (GDPR)[46]. Accordingly, there are many regulations to achieve the goals. In Recital 69 of the AI Act, the right to privacy and protection of personal data must be guaranteed throughout the entire lifecycle of the AI system. In this regard, the principles of data minimisation and data protection by design and by default, as set out in Union data protection law, are applicable when personal data are processed. Moreover, providers may adopt measures to minimize risks to personal data, such as anonymisation and encryption. Technologies can also be used to bring algorithms to the data, enabling AI training without transmitting or copying raw data between parties.

Especially, Article 10 of this Act strictly regulates data governance. High-risk AI systems must meet quality criteria whenever data sets are used. Data sets for training, validation, and testing must have high-quality data which is supported by appropriate data governance and management practices. Besides, to facilitate compliance with the Union Data Protection Law, it is important to ensure transparency regarding the original purpose of personal data collection.[47] One of the practices that shall be of particular concern is data collection processes, the origin of data, the original purpose of the personal data collection, and examination in view of possible biases that may have a negative impact on fundamental rights or lead to discrimination. In certain special cases, to examine, detect, prevent and mitigate possible biases (Art.10(2)(f).(g) of the AI Act), providers of these systems may exceptionally process special categories of personal data, provided that appropriate safeguards for the fundamental rights and freedoms of individuals are in place, as outlined in Article 10.5 of the AI Act. These safeguards include: ensuring that special categories of personal data are subject to technical limitations on their reuse and are protected by state-of-the-art security and privacy-preserving measures, including pseudonymisation; or implementing measures to ensure the personal data

[45] United Nations Human Rights Office of the High Commissioner, *Taxonomy of Human Rights Risks Connected to Generative AI* https://www.ohchr.org/sites/default/files/documents/issues/business/b-tech/taxonomy-GenAI-Human-Rights-Harms.pdf accessed 1 March 2025.

[46] The aspect of the protection of personal data is safeguarded in particular by Regulations (EU) 2016/67946, (EU) 2018/172546 of the European Parliament, and the AI Act, which provide the basis for sustainable and responsible data processing, including where data sets include a mix of personal and non-personal data.

[47] Bence Kis Kelemen and Balázs Hohmann, 'A Schrems ítélet hatásai az európai uniós és magyar adattovábbítási gyakorlatokra (The Effects of the Schrems Judgment on EU and Hungarian Data Transfer Practices)' (2016) 2–3 *Infokommunikáció és Jog* 64-68.

processed are secured, protected, and subject to suitable safeguards, such as strict controls and documentation of access, to prevent misuse and ensure that only authorised personnel have access to the data, with appropriate confidentiality obligations. So, the purpose limitation binds the data controller to a purpose - the original purpose for which the data was first processed under their control.[48] Besides, it also restricts the use of data: it is not legal to process data without a purpose and outside of the purposes for which the data was processed first.[49]

Secondly, AI systems can affect equality and non-discrimination. The international human rights framework grants all people the right to equal protection against discrimination. The UDHR declared that "*Everyone is entitled to all the rights and freedoms set forth in this Declaration, without distinction of any kind, such as race, colour, sex, language, religion, political or other opinion, national or social origin, property, birth or other status."* (Article 2); And *"All are equal before the law and are entitled without any discrimination to equal protection of the law. All are entitled to equal protection against any discrimination in violation of this Declaration and against any incitement to such discrimination."* (Article 7).

AI-driven applications can pose significant dangers because the AI systems may be exploited to reinforce inequality and discrimination. In some cases, AI systems make errors or are purposely utilized as a means to produce outputs that strengthen biases. These facilitate various forms of discrimination in society. For example, the incident involving Microsoft occured when they released an application on Twitter.[50] That incident violated human dignity on a large scale and affected a large number of users within a short amount of time. Another case is Amazon, which found that its algorithm used for hiring employees was biased against women. The reason for that algorithm was based on the number of resumes submitted over the past ten years, and since most of the applicants were men, it was trained to favor men over women.[51]

Therefore, in all circumstances, the risks of discrimination must be prevented and mitigated, particularly for groups that are at a higher risk, including

[48] Indra Spiecker genannt Döhmann, 'AI and Data Protection' in Larry A Dimatteo, Cristina Poncibò and Michel Cannarsa (eds), *The Cambridge Handbook of Artificial Intelligence: Global Perspectives on Law and Ethics* (Cambridge University Press 2022) 132. https://doi.org/10.1017/9781009072168.015

[49] Article 5(1) and 6(1) of the GDPR

[50] On March 23, 2016, Microsoft released Tay to the public on Twitter. Tay is designed to mimic a stereotypical high school girl. Microsoft hoped that Tay would discover patterns in language through "her" interactions with Twitter users and then utilize similar patterns to create her own tweets. At first, Tay engaged harmlessly with her growing number of followers with banter and lame jokes. However after only a few hours, Tay started tweeting highly offensive things. Within 16 hours of her release, Tay posted over 95,000 tweets, a troubling number of which were abusive and offensive.
See it: Oscar Schwartz, 'Microsoft's Racist Chatbot Revealed the Dangers of Online Conversation' (2016) https://spectrum.ieee.org/in-2016-microsofts-racist-chatbot-revealed-the-dangers-of-online-conversation accessed 1 March 2025.

[51] Jeffrey Dastin, 'Insight - Amazon Scraps Secret AI Recruiting Tool That Showed Bias Against Women' (Reuters, 2018) https://www.reuters.com/article/world/insight-amazon-scraps-secret-ai-recruiting-tool-that-showed-bias-against-women-idUSKCN1MK0AG accessed 1 March 2025.

women, children, older adults, economically disadvantaged individuals, members of the LGBT community, persons with disabilities, and racial, ethnic, or religious groups[52]. To achieve this aim, some regulations were enacted and need to be strictly applied:

First, related to the data sets, Recital 67 addresses the management of input data to mitigate possible biases in the data sets. If there are biases in underlying data sets, the outputs could be influenced and thereby perpetuate and amplify existing discrimination. Thus, the data sets should be as complete and free of errors as possible. This requirement should not affect the use of privacy-preserving techniques in the context of the development and testing of AI systems. The requirements related to data governance, such as verification of data governance, data set integrity, and data training, validation and testing practices, can be fulfilled by relying on third parties that offer certified compliance services.

Second, according to Recital 80, as part of implementing the United Nations Convention on the Rights of Persons with Disabilities, persons with disabilities are ensured the right to access and use AI systems on an equal basis with others. Nowadays, AI systems are becoming increasingly important. Thus, the application of universal design principles to all new technologies and services should ensure that everyone potentially affected by AI technologies, including persons with disabilities, is protected in their inherent dignity and diversity. This is one of the indications of equality and non-discrimination. As a result, providers need to meet the requirements of ensuring full compliance with these requirements by design. In particular, the necessary measures should be integrated into the design of the high-risk AI system wherever possible.

Third, the design and development of AI systems must ensure that natural persons can exercise oversight during the operation period. This method partly prevents and minimises the risks of violating fundamental rights as well as discrimination. The high-risk AI system shall be managed through oversight measures that are commensurate with the risks, level of autonomy, and context of use. According to Article 14(3), high-risk AI systems shall be ensured through one or both of the following types of measures: (i). measures identified and built, when technically feasible, into the high-risk AI system by the provider before it is placed on the market or put into service; (ii). measures identified by the provider before placing the high-risk AI system on the market or putting it into service and that are appropriate to be implemented by the deployer.

Moreover, Recital 27 focuses on the connection between the AI Act and seven principles of the 2019 Ethics guidelines for trustworthy AI, developed by the independent AI HLEG appointed by the Commission. Among these principles, related to diversity, non-discrimination, and fairness, all stakeholders, including

[52] Council of Europe Commissioner for Human Rights (n 2).

industry, academia, civil society, and standardisation organisations, are encouraged to develop AI systems to enhance this standard. These systems should diversify actors and promote equal access, gender equality, and cultural diversity, while avoiding discriminatory impacts and unfair biases that are prohibited by Union or national law.

The AI Act also addresses AI systems identifying or inferring the emotions or intentions of individuals based on biometric data, which may lead to discriminatory outcomes. Thus, the placing on the market, development, or use of these systems for detecting emotional states in the workplace and education environment should be prohibited. Furthermore, according to Recital 56, AI systems used in education or vocational training, in particular for determining access or admission, for assigning persons to educational and vocational training institutions or programmes at all levels, for evaluating learning outcomes of persons, for assessing the appropriate level of education for an individual, must be categorized as high-risk systems, as they may pose discrimination based on historical patterns, such as against women, certain age groups, persons with disabilities, or persons of certain racial or ethnic origins or sexual orientation. It is a significant danger since the educational and professional course of a person's life may be affected, especially their ability to secure a livelihood.

On the other hand, public and Union funding is allocated to support and promote research and development of AI solutions that contribute to socially and environmentally beneficial outcomes, such as AI-based solutions to enhance accessibility for persons with disabilities and address socio-economic inequalities. In addition, projects should be based on the principle of interdisciplinary cooperation between AI developers, experts on inequality and non-discrimination, accessibility, consumer rights, environmental rights, digital rights, and academics to enhance efficiency.

However, AI systems used by law enforcement agencies, such as the police, carry the particular risk of significant interference with individual rights. In particular, the State may misuse or abuse AI systems for religious, ethnic, racial persecution or political opinion. For example, real-time facial recognition technology (FRT) might be run on footage from live CCTV systems. Police in Moscow have used FRT to detain a girl preventively, holding her for a few hours for protesting Russia's war in Ukraine.[53] Moreover, the accuracy of FRT is known to be unequally distributed among different demographic groups and biased against already marginalized populations.[54] Several studies have shown that FRTs carry

[53] Darren Loucaides, 'The Changing Face of Protest' (2024) https://restofworld.org/2024/facial-recognition-government-protest-surveillance/#/an-end-to-privacy accessed 1 March 2025.
[54] Lukas Arnold, 'How the European Union's AI Act Provides Insufficient Protection Against Police Discrimination' (2024) https://www.law.upenn.edu/live/news/16742-how-the-european-unions-ai-act-provides accessed 1 March 2025.

racial bias. Misidentification by AI systems also adds to the risk that a government could abuse AI systems to silence dissent and persecute opponents.

Thirdly, *AI systems can affect freedom of expression and access to information rights*. The right to freedom of expression is both a fundamental human right in itself and is also core to the exercise of other rights. The UDHR declared that *everyone has the right to freedom of opinion and expression; this right includes freedom to hold opinions without interference and to seek, receive and impart information and ideas through any media and regardless of frontiers* (Article 19).

AI systems may seriously impact access to, search for, and sharing of information. In fact, the scope of access to information is a key element of the right to freedom of expression.[55] A variety of AI systems are designed to search for information as well as to support study and work (such as ChatGPT, Claude AI, etc.), which are becoming increasingly popular. Meanwhile, generative AI can create false content or misleading content that appears human-generated and authoritative at scale and may pose risks to the right to freedom of expression in various ways.[56] On the other hand, if people are prevented from accessing information, this may violate their rights to freedom of expression and access to information.

Therefore, the State has the responsibility for creating a diverse and pluralistic information environment. Concurrently, AI-driven content may have negative effects on freedom of expression, access to information, and freedom of opinion. Therefore, it must be strictly controlled. One of the fundamental requirements is to ensure reliable data sources, comply with Union Law on copyright, related rights, and authenticity standards. According to Article 13, the provider and deployer must ensure transparency and fulfill their obligation to provide information. High-risk AI systems must ensure transparent operation to allow deployers to interpret the system's output appropriately. This Article also requires that systems have instructions for use, either in an appropriate digital format or another accessible form with concise, complete, correct, and clear information that deployers can easily access and understand.[57] In particular, the characteristics and capabilities of the high-risk AI system can provide information to explain its output. Moreover, it must include the identity and contact details of the provider and, where applicable, of its authorized representative. In Article 16, one of the obligations of providers of high-risk AI systems is to indicate on the high-risk AI system or, where that is not possible, on its packaging or its

[55] Irene Khan, *Disinformation and Freedom of Opinion and Expression: Report of the Special Rapporteur on the Promotion and Protection of the Right to Freedom of Opinion and Expression* (UN Doc A/76/258, 2021) https://documents-dds-ny.un.org/doc/UNDOC/GEN/G21/085/64/PDF/G2108564.pdf accessed 1 March 2025.

[56] United Nations Human Rights Office of the High Commissioner (n 61).

[57] Balázs Hohmann, Adrián Fábián and Gergely László Szőke, 'The Shades of the Concept of Transparency on the Horizon of European Technology Law and Platform Regulation' (2025) 15(1) *Juridical Tribune* 53-54. https://doi.org/10.62768/TBJ/2025/15/1/03

accompanying documentation, as applicable, their name, registered trade name or registered trade mark, the address at which they can be contacted. These tools enable individuals to verify the source of information, helping them to avoid manipulation and misinformation.

On the other hand, according to Recital 134, besides requiring transparency obligations for AI-generated content, the regulation respects the freedom to express ideas creatively. The transparency requirement does not hinder the enjoyment, display, or exploitation of the work. In other words, compliance with the transparency obligation is required, provided that the use of the AI system or its output does not infringe upon the right to freedom of expression and the right to freedom of the arts and sciences. Through this regulation, the rights to access information and freedom of expression are upheld.

Last but not least, Article 5(1)(a) regulates the prohibition of AI practices that intentionally employ manipulative or deceptive techniques. These AI systems significantly distort individuals' ability to make informed decisions. As a result, it is possible that the right to freedom of expression can be seriously infringed upon. Natural persons may be manipulated by misinformation from AI sources, leading to misunderstanding or holding a biased view of an issue. Thus, the regulation to prohibit this type of AI Act proves to be highly protective of this human right.

AI systems can affect the right to work and to gain a living. It is an essential human right as it can decide the livelihood of a person in particular and even that of a family in general. It cannot be denied that AI tools play a significant role in enhancing labor productivity and replacing humans in hazardous work environments. However, the right to work can be seriously affected by the development of AI, as well as drastically altering economic and labor markets. As the capability of AI to accelerate automation increases, various types of jobs may be replaced by AI. As a result, it reduces the amount of labor in the market. For example, companies and factories may replace workers with AI tools or pause hiring for roles that may be performed by generative AI in the future.

In the meantime, in Article 23 of UDHR, everyone has the right to work, to free choice of employment, to just and favourable conditions of work and to protection against unemployment. And everyone who works has the right to just and favourable remuneration ensuring for himself and his family an existence worthy of human dignity, and supplemented, if necessary, by other means of social protection.

In the AI Act, AI systems used in employment, workers' management, and access to self-employment are classified as high-risk. Especially under Annex III, Section 4 clarifies two groups, including (i). AI systems are intended to be used for the recruitment or selection of natural persons, in particular to place targeted job advertisements, to analyse and filter job applications, and to evaluate candidates. (ii). AI systems are intended to be used to make decisions affecting terms of work-related relationships, the promotion or termination of work-related contractual

relationships, to allocate tasks based on individual behaviour or personal traits or characteristics, or to monitor and evaluate the performance and behaviour of persons in such relationships. These AI systems directly affect employment and promotion opportunities. These are also used to evaluate the performance of employees in the workplace and even to terminate them. Thus, they must be strictly regulated. According to Article 9 of the AI Act, high-risk AI systems must have risk management systems - a continuous iterative process planned and run throughout the entire lifecycle of a high-risk AI system, which aims to eliminate or reduce risks related to the use of high-risk AI systems.

Moreover, AI systems underwent rapid development during this period, which poses potential risks to the labor market. Therefore, the authorities need to frequently track and evaluate trends in the labor market, such as the number and types of jobs created and lost due to AI; give suggestions on methods to improve personal skills and knowledge, as well as invest in developing a high-quality labor market. The government should update education curricula to ensure access to jobs requiring competencies related to AI systems. Focus on training the workforce in 'low automation' job sectors to mitigate job losses or unemployment. Reassigning jobs is also necessary to protect the right to work and to gain a living. Thereby fully exploiting the advantages while safeguarding human rights.

In short, the development of technology has had a prolonged effect on human rights, starting from the Industrial Revolution in general and the introduction of artificial intelligence in particular. Most of the impacts are undoubtedly positive and productive in nature.[58] However, the negative effects could become colossally unprecedented if the state does not implement and enforce appropriate regulations. The AI Act and relevant regulations are an important step in the process of protecting human rights. During this period, the protection of human rights is not a challenge for any single country; rather, it is a multinational issue.

[58] Somesh Sankhala and Falguni Mundhra, 'Artificial Intelligence vs Human Rights' (2023) 3(2) *Jus Corpus Law Journal* 401.