

*Marta Grabowska**

Artificial Intelligence Systems in the Context of UN Resolution A/78/L.49, Council of Europe Framework Convention CETS No. 225, and EU Regulation 2024/1689

Abstract

The aim of this article is to examine the approach to artificial intelligence reflected in three key international instruments adopted in 2024: the UN Resolution A/78/L.49, the Council of Europe Framework Convention CETS No. 225, and the European Union Regulation 2024/1689. Although these texts are commonly referred to as legal acts on artificial intelligence, their actual subject matter is more precisely defined as artificial intelligence systems – that is, specific entities with which humans interact. Despite differences in content and legal force, each document delineates, in its own way, the role of AI systems within human society and establishes safeguards against their undue influence. Additionally, each assigns responsibilities aligned with the objectives of the respective issuing organisation. A shared point of reference among them is the framework of international human rights protection, supplemented by various additional requirements unique to each instrument. This article analyses the substantive provisions of the three documents, comparing their terminology, scope, proposed protective mechanisms, and procedures for implementation.

Keywords: Artificial Intelligence, Artificial Intelligence Systems, UN Resolution A/78/L.49, Council of Europe Framework Convention CETS No. 225, European Union Regulation 2024/1689, Human Rights

* **Marta Grabowska** – University of Warsaw, Poland, e-mail: mgrabowska@uw.edu.pl, ORCID ID: 0000-0002-4481-0545.

Introduction

The *PWN Dictionary of the Polish Language* (*Słownik Języka Polskiego PWN*) provides the following definition of artificial intelligence: “artificial intelligence – a branch of computer science that studies the rules governing human mental behaviour and creates computer programmes or systems that simulate human thinking” (*Słownik PWN*, 2025).

The origins of artificial intelligence can be traced to the work of the eminent British mathematician Alan Turing (1912–1954). Following the Second World War – during which he was engaged in decrypting the codes of the German ENIGMA machine (Hodges, 2014) – Turing conducted the first so-called “Turing Test” in 1950. This experiment aimed to determine whether a machine could engage in conversation in a manner that demonstrated human-like cognitive abilities (Oppy, 2021). Although Turing’s untimely death curtailed further development of his research, the promising results of his early work inspired a gathering of leading scientists at Dartmouth College in Hanover, New Hampshire, in 1956. This meeting is widely regarded as the formal beginning of artificial intelligence research. The participants posited 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” (Veisdal, 2019). John McCarthy, a professor of mathematics and the organiser of the Dartmouth Conference, was the first to identify artificial intelligence as a distinct field of research (Schneppat, 2025). The development of artificial intelligence is commonly divided into three stages: (1) the logic-based approach, (2) the knowledge-based expert systems approach, and (3) the data-driven approach (Flasiński, 2016).

The first stage, emerging in the 1950s and 1960s, resembled the early Turing Tests and relied on logical operations applied to simple textual inputs, constrained by the limited computing power of the time. The second stage, which gained prominence in the 1970s and 1980s, introduced expert systems capable of handling more complex tasks. However, these systems lacked the ability to learn and adapt. It was not until the early 21st century that a technological breakthrough occurred, driven by a dramatic increase in computational power, the availability of vast datasets, the emergence of large-scale computing and cloud technologies, and significant advances in related disciplines such as neuroscience and cognitive science.

This development paved the way for the adoption of neural networks, which enabled the simulation of human cognitive and learning processes. Given the rapid advancement of systems that utilise artificial intelligence to replicate human actions, there is an increasing need to

regulate their impact to ensure they do not pose a threat to humanity. While various recommendations and legal frameworks have been or are being developed at the national, corporate, and bilateral levels, this article focuses on three key international documents adopted in 2024 that are crucial for shaping the future of artificial intelligence in human society. These include: the UN Resolution A/78/L.49 (United Nations, 2024a), the Council of Europe Framework Convention CETS No. 225 (Council, 2024b), and the European Union Regulation 2024/1689 (Regulation, 2024b).

UN Resolution A/78/L.49

If arranged chronologically, the first of these documents is the United Nations Resolution entitled *Seizing the opportunities of safe, secure and trustworthy artificial intelligence systems for sustainable development*, adopted by the UN General Assembly on 21st March 2024 (A/78/L.49) (United Nations, 2024a). This Resolution, proposed by the United States and supported by 123 UN Member States, including China and Russia, was adopted by consensus (Mishra, 2024; United Nations, 2024b). Its primary objective is to oversee artificial intelligence systems with a focus on safeguarding human rights – particularly the right to privacy – while harnessing AI’s potential to implement the 17 Sustainable Development Goals (SDGs) outlined in the UN’s *2030 Agenda for Sustainable Development* (A/RES/70/1) (United Nations, 2015b), across all economic, social, and environmental dimensions. Adopted by 193 countries in 2015, the *Agenda* sets a target for implementation by 2030.

The Resolution on artificial intelligence was published in all official UN languages: English (being the official version), Arabic, Chinese, French, Spanish, and Russian. Although the Resolution has a global scope and its adoption can undoubtedly be regarded as a historic event, it is not a legally binding document.

Akin to other UN resolutions, it serves as an appeal, manifesto, or proclamation, as only resolutions of the UN Security Council carry binding legal force. The Resolution consists of a preamble and an operative part. The preamble references key UN documents, including the *Charter of the United Nations* (Charter, 1945) and the *Universal Declaration of Human Rights* (United Nations, 2025b), alongside other documents specifically related to information and communication technologies.

While the Resolution does not provide a definition of artificial intelligence itself, the preamble includes a definition of artificial intelligence systems, which states: “Recognizing that safe, secure and

trustworthy artificial intelligence systems – which, for the purpose of this resolution, refers to artificial intelligence systems in the non-military domain, whose life cycle includes the stages: pre-design, design, development, evaluation, testing, deployment, use, sale, procurement, operation and decommissioning – are such that they are human-centric, reliable, explainable, ethical, inclusive, in full respect, promotion and protection of human rights and international law, privacy-preserving, sustainable development-oriented, and responsible – [and] have the potential to accelerate and enable progress towards the achievement of all 17 Sustainable Development Goals and sustainable development in its three dimensions – economic, social and environmental – in a balanced and integrated manner; promote digital transformation; promote peace; overcome digital divides between and within countries; and promote and protect the enjoyment of human rights and fundamental freedoms for all, while keeping the human person at the centre” (pp. 2–8).

The title and content of the preamble clearly indicate that the central theme of this Resolution is its role as a vital tool for advancing the *2030 Agenda for Sustainable Development* (United Nations, 2015b), which encompasses 17 goals aimed at improving socio-economic and environmental conditions for humanity. These goals seek to eliminate poverty and hunger; ensure good health and quality of life; provide quality education; achieve gender equality; ensure clean water and sanitation; offer clean and affordable energy; promote decent work and economic growth; foster innovation in industry and infrastructure; reduce inequalities; build sustainable cities and communities; encourage responsible production and consumption; address climate change; protect life on land and under water; promote peace, justice, and strong institutions; and foster partnerships to achieve all these objectives (United Nations, 2025a).

The preamble’s references to key documents further confirm the global perspective on the role of artificial intelligence systems in achieving these goals. Notable references include: *Transforming Our World: the 2030 Agenda for Sustainable Development* (A/RES/70/1) (United Nations, 2015b); *Impact of Rapid Technological Change on the Achievement of the Sustainable Development Goals and Targets* (A/RES/77/320) (United Nations, 2021); *Information and Communications Technologies for Sustainable Development* (A/RES/78/132) (United Nations, 2023a); and *Science, Technology and Innovation for Sustainable Development* (A/RES/78/160) (United Nations, 2023b).

The preamble also highlights the efforts of the International Telecommunication Union (ITU), undertaken in collaboration with

40 other UN agencies and the Swiss Government, which led to the establishment of the *AI for Good* platform in 2017 (ITU, 2025).

This initiative serves as a catalyst for leveraging the development of artificial intelligence systems in support of the 17 Sustainable Development Goals. Launched during the *AI for Good Global Summit* held in Geneva in 2017, the platform brings together leading experts and stakeholders from diverse sectors – including industry, academia, international organisations, governments, and non-governmental organisations – with the shared objective of promoting AI systems that contribute to sustainable development. It is now a vibrant initiative with over 37,000 active participants representing more than 180 countries, meeting every year in Geneva and other parts of the world. The platform also encourages young people to cooperate and intensifies efforts in response to the postulates of the G20, the UN’s High-Level Advisory Body on Artificial Intelligence (United Nations, 2025c), and the UN’s *Global Digital Compact* (United Nations, 2024c) in the field of artificial intelligence. It further seeks to improve the governance of artificial intelligence systems for sustainable development, aiming to ensure their use for the common good. With only five years remaining until the implementation deadline of the *2030 Agenda*, artificial intelligence systems are increasingly viewed as essential tools to support progress toward its goals.

The operative part of the Resolution, comprising 13 points, outlines a range of recommendations, including calls for the participation of all countries – particularly developing nations – as well as other relevant stakeholders in the creation of safe, ethically sound, and inclusive AI systems accessible to all. The Resolution emphasises the need to develop the infrastructure necessary for the deployment of AI technologies, increase financial investments in AI development and innovation, and ensure that such progress aligns with the protection of privacy and other fundamental human rights. Additional recommendations include efforts to improve public skills and digital literacy for interacting with AI systems, the promotion of responsible data governance given the central role of data in AI, and the formulation of legal frameworks at all levels that reflect the values articulated in core UN documents – namely, the *United Nations Charter* (Charter, 1945), the *Universal Declaration of Human Rights* (United Nations, 2025b), and the *2030 Agenda for Sustainable Development* (United Nations, 2015b).

The adoption of this Resolution by the UN General Assembly resonated widely across the globe. Its central appeal to humanity can be succinctly captured in the call: “Let us – humans – govern this technology rather than have it govern us”.

Council of Europe Framework Convention CETS No. 225

The Council of Europe (Council, 2025) is an international governmental organisation dedicated to promoting human rights, democracy, and cooperation among its Member States, particularly in the realm of culture. Established on 5th May 1949 in London by 10 founding members, the organisation now includes nearly all European countries, as well as several non-European states. Headquartered in Strasbourg, France, the Council of Europe currently comprises 46 Member States, collectively representing a population of approximately 700 million people. Its primary objectives include, but are not limited to, the defence of fundamental rights and freedoms, such as freedom of speech, the right to free and fair elections, freedom of the media, social rights, privacy, and the protection of national minorities and their languages, the rights of people with disabilities, as well as addressing violence against women and children, corruption, cybercrime, money laundering, and other societal threats. The Council of Europe also seeks to foster civil society and addresses contemporary human rights issues, such as bioethics, environmental threats, and artificial intelligence.

It operates through binding international legal instruments known as conventions or treaties, which encompass a broad range of human rights concerns. Notably, it remains the only organisation of its kind globally. These conventions are originally published in two official languages: English and French. They are adopted by the Committee of Ministers of the Council of Europe – following deliberation within the Parliamentary Assembly, the organisation's decision-making body composed of the Ministers of Foreign Affairs of the Member States – and subsequently signed and ratified by the relevant parties. The entry into force of a convention depends on the number of states that ratify the instrument. Ratification of conventions is voluntary for the parties, with the exception of the founding convention, which is mandatory for all Member States (Council, 1950). Both Member and non-Member States of the Council of Europe, as well as international organisations, may ratify individual conventions. This approach facilitates the extension of the European *acquis* in the field of human rights to other regions of the world. In 2024, the Committee of Ministers of the Council of Europe adopted its 225th convention: the *Council of Europe Framework Convention on Artificial Intelligence and Human Rights, Democracy and the Rule of Law* CETS No. 225 (Council, 2024b). The Convention was formally adopted on 17th May 2024 and presented to the interested parties for further processing in Vilnius on 5th September 2024. It will enter into force once it has been ratified by

at least five parties, including a minimum of three member states of the Council of Europe. To date, the Convention has not yet been ratified by any party. However, it has been provisionally signed by ten member states – including Norway and the United Kingdom – two non-member states, including the United States of America, and one international organisation, namely, the European Union. As a result, the Convention has not yet entered into force. The ratification process involves the incorporation of the convention’s provisions into the domestic legal framework of the party concerned. Nevertheless, its adoption by the Committee of Ministers – the Council of Europe’s principal decision-making body – and its submission for further processing represent a significant milestone in the global effort to regulate the relationship between humans and artificial intelligence. The explanatory report accompanying the Convention (Council, 2024a) emphasises that its primary focus is the protection of human rights, democracy, and the rule of law, as reflected in its title. It explicitly states, however, that the Convention does not aim to regulate the economic or market dimensions of artificial intelligence systems. The preamble to the Convention underscores the expectation that incorporating its provisions into the domestic legal frameworks of the parties will ensure adherence to the established standards, thereby promoting safety, social well-being, and respect for human dignity.

The main body of the Convention comprises eight chapters. Chapter I outlines the object and purpose of the Convention. As in the previously discussed international instruments, the term “artificial intelligence system” is employed, with the Convention specifying that its scope encompasses the entire life cycle of such systems. It is emphasised that this life cycle must be aligned with the principles of human rights, democracy, and the rule of law. Parties are required to implement legal frameworks that are consistent with the Convention’s provisions in a manner that most effectively safeguards these values. Moreover, the method by which each party intends to implement the Convention must be submitted to the Council of Europe for approval during the process of signature or ratification. The Convention also explicitly states that its provisions do not apply to systems used in scientific research or within the defence sector. For the purposes of this Convention, an artificial intelligence system is defined thusly “artificial intelligence system – a machine-based system that, for explicit or implicit objectives, infers, from the input it receives, how to generate outputs such as predictions, content, recommendations or decisions that may influence physical or virtual environments. Different artificial intelligence systems vary in their levels of autonomy and adaptiveness after deployment” (Art. 2).

Chapter II of the Convention sets out provisions concerning the protection of human rights, which must be upheld in the domestic legislation of parties implementing the Convention. These provisions are required to align with both international and national legal frameworks. Furthermore, artificial intelligence systems must not undermine democratic processes or the rule of law. The Convention also mandates the establishment of mechanisms to assess the effectiveness of the actions taken.

Chapter III addresses the activity of artificial intelligence systems throughout their lifecycle and outlines a set of requirements that each party implementing the Convention is obliged to adopt. These requirements pertain to:

1. human dignity and individual autonomy,
2. transparency and oversight,
3. accountability and responsibility,
4. equality and non-discrimination,
5. privacy and personal data protection,
6. reliability,
7. safe innovation.

Chapter IV sets out provisions concerning legal remedies applicable throughout the entire lifecycle of artificial intelligence systems. It emphasises the necessity of access to relevant documentation and information regarding potential risks posed by such systems. The Convention underscores the importance of safeguarding the right to object to the use of artificial intelligence systems and to challenge decisions made through their application. It also mandates the establishment of mechanisms for lodging complaints with competent authorities and stresses the need for transparent and accessible procedures at both national and international levels. Additionally, it requires that individuals be informed when they are interacting with a system rather than a human being.

Chapter V is dedicated to the principles for risk assessment and the mitigation of adverse outcomes in situations where human rights, democracy, or the rule of law may be endangered.

Chapter VI concerns the procedures for implementing the Convention. It emphasises the principle of equal treatment for all future users, including persons with disabilities and children. The chapter highlights the importance of public consultation processes, the promotion of adequate digital literacy, and the provision of safeguards to ensure the protection of human rights in line with their established formulations.

Chapter VII sets out the requirements for verification and cooperation processes. It establishes a Conference of the Parties, assisted by the

General Secretariat of the Council of Europe, as a forum for information exchange regarding the Convention and for the periodic review of its implementation. Each party is required to submit an initial report within two years of acceding to the Convention and to provide subsequent reports at regular intervals. Furthermore, each party commits to establishing at least one mechanism to monitor compliance with the Convention's provisions. The chapter also introduces a framework for international cooperation aimed at facilitating the exchange of information on both the positive and negative impacts of artificial intelligence systems, and at supporting non-party entities in addressing issues falling within the Convention's scope.

Chapter VIII comprises the final provisions of the Convention. These relate, among other things, to procedures for submitting amendments to the Convention's text, as well as for its signing and ratification. Notably, Article 32, titled "Territorial Application", introduces a potentially contentious issue. It stipulates that each party to the Convention may specify the internal territory within which the Convention's provisions will apply, and may subsequently alter this territorial scope. This diverges significantly from the model used by the European Union, where regulations are addressed to Member State governments and are uniformly applicable across the entirety of a Member State's territory. In contrast, under this Convention, a party – whether a state or an international organisation – may determine the geographical area within which the Convention will be implemented, or even revise this scope at a later date. This raises concerns, particularly in scenarios where more authoritarian governments might seek to limit the Convention's application. The question may even be posed, albeit provocatively, whether it would be legally permissible for a state such as Poland to implement the Convention solely within the city of Kraków, while excluding the remainder of its territory. The situation becomes even more complex in the context of countries with federal systems. Conventions such as this one typically do not prescribe penalties for non-compliance. However, they usually indicate that each party, during the ratification process, should adopt appropriate criminal law instruments to ensure the effective implementation of the Convention.

The timeline for the Convention's entry into force remains uncertain, as there is no obligatory deadline for ratification and the process is entirely voluntary. Although the protections offered by the Council of Europe Framework Convention CETS No. 225 (Council, 2024b) are of pressing importance, it is currently impossible to predict when the Convention will attain binding force. Ratification must be undertaken individually by each party, meaning that the ratification

of the Convention by the European Union does not imply automatic ratification by its Member States. To date, no EU Member State has signed the Convention (a particularly striking fact), not even Lithuania, where the ceremony marking the opening of the Convention for signature was held. Unfortunately, this process may take a very long time, judging by the course of events surrounding, for example, the Convention on Cybercrime (ETS No. 185) (Council, 2001), adopted on 23rd November 2001 in Budapest. It entered into force only on 1st July 2004, and Poland ratified this document on 1st June 2015 – 14 years later! One explanation for such delays is the need to adequately prepare the justice system for the implementation of the Convention's provisions, including equipping law enforcement agencies and courts to handle relevant cases effectively and streamlining procedures for filing complaints with the European Court of Human Rights in Strasbourg (European, 2025), which serves as the final and authoritative judicial instance. This was, for instance, the justification provided by the Polish authorities for the delayed ratification of the Convention on Cybercrime. As for when CETS Convention No. 225 (Council, 2024b) becomes a fully operational protective instrument – only time will tell.

The European Union's Regulation (2024/1689)

The next document is Regulation (EU) 2024/1689 of the European Parliament and of the Council, adopted on 13th June 2024, commonly referred to as the Artificial Intelligence Act. This regulation is binding in its entirety and directly applicable in all Member States of the European Union. It is also relevant to the countries of the European Economic Area. The primary objective of the regulation is to structure and regulate the functioning of artificial intelligence systems within the single European market (Regulation, 2024b).

The origins of the AI Act can be traced back to 2017, when the European Parliament adopted the Resolution of 16th February 2017 with recommendations to the Commission on civil law rules on robotics (2015/2103(INL)) (European, 2017). This document outlined general principles concerning the development of robotics and artificial intelligence for civil use. Subsequently, on 25th April 2018, the European Commission issued a Communication from the Commission to the European Parliament, the European Council, the European Economic and Social Committee, and the Committee of the Regions – *Artificial Intelligence for Europe* (COM(2018) 237 final) (European, 2018), which marked the formal launch of a European initiative on artificial intelligence.

The objectives of this initiative included the implementation of AI systems across the EU economy to enhance its economic and industrial capacity, prepare for the socio-economic transformations brought about by AI, and ensure the development of an ethical and legal framework aligned with the values enshrined in the Charter of Fundamental Rights of the European Union (*Charter*, 2012). Among the key measures outlined in this instrument were increased financial investments in research, innovation, and development for the period 2018–2020, as well as the creation of an “AI on Demand” platform. This platform was intended to serve as a collaborative laboratory for all interested stakeholders within the Member States, with a particular emphasis on supporting small and medium-sized enterprises. Developed between 2019 and 2021 under the Horizon 2020 programme, the platform is currently used to facilitate the sharing of resources, tools, and knowledge related to artificial intelligence algorithms (AI4EU – AI on Demand) (AI on Demand, 2025). In June 2018, the European Commission appointed the High-Level Expert Group on Artificial Intelligence (AI HLEG) (European, 2024a), tasked with providing advisory services to the Commission. The group comprised 52 specialists from diverse scientific backgrounds, including computer science, robotics, philosophy, and law, alongside representatives from business and social sectors. In the Commission Communication of 8th April 2019, entitled *Building Trust in Human-Centric Artificial Intelligence* (COM(2019)168) (European, 2019), the European Commission outlined seven key requirements developed by AI HLEG that must be fulfilled for the use of artificial intelligence systems to be considered trustworthy. These requirements are: human agency and oversight; technical robustness and safety; privacy and data governance; transparency; diversity, non-discrimination, and fairness; social and environmental well-being; and accountability. Although the AI HLEG concluded its mandate in July 2020, the forum established around it – comprising hundreds of participants organised within the AI Alliance (European, 2024b) – continues its activities in cooperation with the European Commission. One of the main outcomes of these efforts was the preparation of the *White Paper on Artificial Intelligence* (COM(2020)65 final/2), published on 19th February 2020 (European, 2020a). The White Paper sets out a policy and regulatory framework for the development of AI systems within the European Union, aiming to establish an “ecosystem of excellence” and an “ecosystem of trust” grounded in European values, with particular emphasis on high-risk AI systems. On the same day, the European Commission also issued the *Communication European Strategy for Data* (COM(2020)66) (European, 2020d), dedicated to the role of data in the development of artificial intelligence. This document outlined a vision for

a data-driven economy and addressed the challenge of regulating access to data in a manner that respects privacy, while also defining the framework for the use of data from both the public and private sectors.

The initiatives described above took place within the framework of the Europe 2020 strategy (Europe, 2010) and can be regarded as preparatory steps toward their eventual materialisation in the form of binding EU legislation. This process was further advanced under the subsequent development framework of the European Union, namely the Multiannual Financial Framework 2021–2027 (MFF) (Council, 2020) – a strategy formulated under the challenging conditions of the COVID-19 pandemic, which also necessitated the creation of an additional economic stimulus program, the Recovery Plan for Europe (Regulation, 2021).

The Multiannual Financial Framework (MFF) 2021–2027 covers seven main areas:

1. Single market, innovation, and digital economy,
2. Cohesion, resilience, and values,
3. Natural resources and the environment,
4. Migration and border management,
5. Security and defence,
6. Neighbourhood and the world,
7. European public administration.

The Recovery Plan for Europe supports only the first three areas of the MFF, which are dominated by two major transformations: the green and digital transitions, both highlighted in Ursula von der Leyen’s address to the European Parliament on 27th November 2019 in Strasbourg (von der Leyen, 2019). The focus on digital transformation – which is also seen as a key tool for achieving climate goals – accelerated the European Commission’s legislative initiatives to structure the digital sphere within the EU, emphasising technology, infrastructure, and data. As part of these efforts, a comprehensive package of regulations has already been adopted, including:

- The Digital Markets Act (2022/1925) (Regulation, 2022a),
- The Digital Services Act (2022/2065) (Regulation, 2022b),
- The European Chips Act (2023/1781) (Regulation, 2023a),
- The Data Act (2023/2854) (Regulation, 2023b),
- The European Digital Identity Framework (2024/1183) (Regulation, 2024a),
- The Artificial Intelligence Act (2024/1689) (Regulation, 2024b).

The latter consists of an extensive preamble with 180 recitals, followed by a main body structured into thirteen chapters and 113 articles, supplemented by thirteen annexes.

Although the title of the Artificial Intelligence Act (Regulation, 2024b) suggests the establishment of harmonised rules for artificial intelligence, the legal text specifically addresses artificial intelligence systems, consistent with the approach outlined in both the previously discussed UN Resolution A/70/L.49 (United Nations, 2024a). and the Council of Europe Convention CETS No. 225 (Council, 2024b). For the purposes of the Regulation, an “AI system” is defined as a “machine-based system designed to operate with varying levels of autonomy, potentially exhibiting adaptiveness after deployment, and, for explicit or implicit objectives, inferring from the input it receives how to generate outputs such as predictions, content, recommendations, or decisions that may influence physical or virtual environments” (Art. 3, par. 1).

Although the “pyramid” model, which classifies artificial intelligence systems into four categories based on the degree of risk (minimal risk, limited risk, high risk, and unacceptable risk), is frequently referenced in various documents (Atack, Limam, 2021), this specific model is not explicitly adopted in the Artificial Intelligence Act. However, the Regulation makes a significant contribution by establishing a list of prohibited AI practices, which outline activities leading to the complete exclusion of certain systems from use. It also sets the criteria for qualifying systems as high-risk AI systems, which can only be deployed under specific conditions.

In the context of the Regulation, “risk” is defined as “the combination of the probability of an occurrence of harm and the severity of that harm” (Art. 3, par. 2). High-risk AI systems, as detailed in Annex III of the Artificial Intelligence Act, include, among others:

- Systems that operate with biometric data,
- Systems that manage critical infrastructure,
- AI systems used in education and vocational training,
- Systems involved in employment procedures,
- AI systems regulating access to private and public services,
- Systems applied in law enforcement,
- AI used in migration, asylum, and border control,
- AI systems involved in the administration of justice and democratic processes.

However, regarding the latter category, Recital 61 clarifies that “The use of AI tools can support the decision-making power of judges or judicial independence, but should not replace it; the final decision-making must remain a human-driven activity”. Furthermore, this list of high-risk categories is subject to potential expansion in the future.

The Artificial Intelligence Act imposes supervisory obligations on manufacturers, distributors, importers, and entities using high-risk AI systems, as well as on state authorities. Each Member State must establish a notifying authority responsible for overseeing the proper designation and functioning of conformity assessment bodies and their subcontractors. The goal is to manage the entire lifecycle of high-risk AI systems through a risk management framework that ensures transparency, accuracy, robustness, and cybersecurity, while also considering the impact on fundamental rights.

Article 6, paragraph 3, clarifies that systems will not be classified as high-risk if they do not pose a significant risk to the health, safety, or fundamental rights of individuals, including not materially influencing decision-making outcomes. The Regulation thus distinguishes between three categories of systems based on the level of risk associated with their use: unacceptable risk; high risk; and systems that do not pose a significant risk.

The Artificial Intelligence Act also establishes requirements for the training, validation, and test data used in AI systems. These principles aim to eliminate gaps, biases, and errors, while ensuring the security of special categories of data, such as personal data. Additionally, the Regulation introduces comprehensive procedures for handling incidents related to AI systems.

A distinct category within the Artificial Intelligence Act is General Purpose AI (GPAI). These are not AI systems in the traditional sense, but rather general models capable of performing a wide range of tasks. These models are trained on vast datasets and only become AI systems when integrated with a user interface or other application layers. GPAI models are typically deployed through libraries, application programming interfaces (APIs), and similar tools. Examples of such models include ChatGPT, DALL·E 2, and Bard, which consist of at least one billion parameters and can execute diverse functions.

Systems that implement these models may pose specific hazards and large-scale risks, including systemic harm. Therefore, if such a system is placed on the market, the provider must notify the European Commission. Based on this notification, the Commission may assess on a case-by-case basis whether the model in question constitutes a systemic risk. The Regulation defines systemic risk as “a risk specific to the high-impact capabilities of general-purpose AI models, having a significant effect on the Union market due to their reach or because of actual or reasonably foreseeable negative consequences on public health, safety, public security, fundamental rights, or society as a whole, which could spread across the

value chain”. In response to these concerns, the European Commission has announced the creation of a dedicated Code of Practice for such systems. Stakeholder participation is encouraged in this process, which is expected to be finalised by May 2025 (European, 2025b).

The Commission is also responsible for publishing a list of General Purpose AI systems. In addition to addressing risks, the AI Act includes provisions designed to foster innovation. These include regulatory sandboxes, which allow the testing of high-risk AI systems under controlled conditions, as well as the possibility of testing outside these sandboxes. The Regulation also provides support for small and medium-sized enterprises (SMEs) to encourage their participation in AI development and innovation.

To ensure effective governance, the AI Act establishes the European Artificial Intelligence Board. This board will be supported by an advisory forum made up of independent scientific experts. The Regulation also introduces a centralised EU database to monitor high-risk AI systems and includes mechanisms for supervising AI systems after their deployment. Non-compliance with the Regulation’s provisions can lead to substantial penalties. Regulation 2024/1689 (Regulation, 2024b) will come into effect on 2nd August 2026, and it will be published in all official languages of the European Union.

Undoubtedly, the Artificial Intelligence Act represents a pioneering regulatory initiative that establishes a general framework for the deployment of AI systems, placing the human being at the centre as the ultimate overseer of all such processes. This approach reflects core European values and underpins the broader objective of achieving European digital sovereignty. The Regulation is intended to govern the functioning of AI systems within the European single market. However, given that many of these systems originate outside the European Union, its provisions are also expected to exert significant extraterritorial influence. Non-EU producers aiming to access the EU market will be required to adapt their products to comply with the Regulation’s standards. Consequently, Regulation 2024/1689 (Regulation, 2024b) is anticipated to become a global benchmark for AI governance.

In Poland, preparatory work is underway to implement the Regulation, including the designation of a national notifying authority and the development of adjudication procedures. The Court of Justice of the European Union in Luxembourg will serve as the final adjudicative body in matters arising from the Regulation.

Artificial Intelligence Systems and Human Rights

As evidenced by the analysis of the aforementioned documents, a fundamental issue lies in the relationship between humans and the “black box”, which constitutes the core of every artificial intelligence system. Typically, individuals lack an understanding of the underlying mechanisms that drive these increasingly complex systems and are often unaware that such mechanisms can be applied to their lives. The documents reviewed reveal that the primary framework for safeguarding human rights in the context of artificial intelligence is grounded in established systems of human rights protection, particularly those of an international nature. These frameworks include the United Nations, the Council of Europe, and the European Union, each of which is supplemented by specific provisions found within the documents discussed. Therefore, creators of artificial intelligence systems intended for use in sectors governed by the aforementioned human rights frameworks should primarily adhere to these frameworks, alongside the additional requirements set forth in the three regulations discussed above. However, it is important to recognise that artificial intelligence systems are not exclusively developed within these regulated areas but are also emerging in other regions of the world. Consequently, the aim of these documents is to establish a protective filter that shields sectors covered by these requirements from technologies that fail to meet these standards, while also serving as a form of supervisory normative authority. This raises the question: Is such a regulatory approach sustainable in the face of global systems such as TikTok or DeepSeek?

Conclusions

In 2024, three pivotal international documents concerning artificial intelligence were adopted: UN Resolution A/78/L.49 (United Nations 2024a), the Council of Europe Framework Convention CETS No. 225 (Council, 2024b), and EU Regulation 2024/1689 (Regulation, 2024b). At present, none of these documents is legally binding; the UN Resolution, by its nature, does not have the force of law, while the Council of Europe Framework Convention and the EU Regulation have yet to enter into force. Notably, none of these documents directly addresses artificial intelligence itself but rather focuses on artificial intelligence systems with which humans interact. This distinction arises from the recognition that artificial intelligence, as a rapidly evolving technology, is not merely a tool but also a distinct field of research. This field encompasses disciplines such as computer science, cognitive science, and philosophy. Humans interact

with specific artificial intelligence systems as products of this rapidly advancing technology, such as diagnostic medical devices, employment qualification systems, unmanned vehicles, and more. It is these systems that necessitate regulatory requirements. However, the documents discussed set different goals for these systems. The UN (United Nations 2024a) views them primarily as tools to accelerate the achievement of the 17 Sustainable Development Goals outlined in the *2030 Agenda for Sustainable Development* (United Nations, 2015b). Recognising that artificial intelligence systems can facilitate more efficient progress towards these goals, the Council of Europe's CETS Convention No. 225 (Council, 2024b) emphasises the expectation that AI systems will contribute to the development of civil society, as well as protect human rights, democracy, and the rule of law. In contrast, the European Union Regulation (2024/1689) (Regulation, 2024b) focuses on regulating the functioning of AI systems within the European market. The range of expectations placed on artificial intelligence systems by these three documents is indeed vast. However, in the areas most impacted by these documents, AI systems continue to be deployed without comprehensive legal protection, unless specific local regulations are in place. The international human rights frameworks of the UN, the Council of Europe, and the European Union serve as the foundation for protecting against the undesirable effects of AI systems, and each of the discussed documents introduces additional requirements. However, not all of these requirements are legally binding at present. Moreover, it remains uncertain whether these requirements can be maintained in light of the technological advancements occurring in other parts of the world. The possibility of deregulatory movements in this field cannot be ruled out. A particularly concerning aspect is that, for instance, the Council of Europe Framework Convention CETS No. 225 (Council, 2024b) has not yet been preliminarily signed by any EU Member State. Whether these regulatory frameworks will be effectively implemented and maintained remains uncertain, and time will ultimately tell how the situation unfolds.

References

- AI on Demand (2025) Available at: <https://www.ai4europe.eu/> (Access 2.09.2025).
- Atack, P. and Limam, A. (2021) *Artificial intelligence: EU sets out „pyramid of risks” to underpin global standards*. Available at: <https://newseu.cgtn.com/news/2021-04-22/EU-s-pyramid-of-risk-aims-to-set-global-standards-for-AI-ZEIADwj8FG/index.html> (Access 3.02.2021).

- Charter (1945) of the United Nations and the Statute of the International Court of Justice. San-Francisco 1945.* Available at: <https://treaties.un.org/doc/Publication/CTC/uncharter.pdf> (Access 2.09.2025).
- Charter (2012) of Fundamental Rights of the European Union.* OJ C 2012/326/, 26.10.
- Council (1950) of Europe, *Convention for the Protection of Human Rights and Fundamental Freedoms* (ETS No. 005). Available at: <https://www.coe.int/en/web/conventions/full-list?module=treaty-detail&treaty-num=005> (Access 2.09.2025).
- Council (2020) *Regulation (EU, Euratom) 2020/2093 of 17 December 2020 laying down the multiannual financial framework for the years 2021 to 2027.* OJ L 433 I/11. 22.12.
- Council (2001) of Europe. *Convention on Cybercrime* (ETS No. 185). Available at: <https://www.coe.int/en/web/conventions/full-list?module=treaty-detail&treaty-num=185> (Access 2.09.2025).
- Council (2024a) of Europe. *Explanatory Report to the Council of Europe Framework Convention on Artificial Intelligence and Human Rights, Democracy and the Rule of Law.* Vilnius. 5.09. Available at: <https://rm.coe.int/1680afae67> (Access 2.09.2025).
- Council (2024b) of Europe. *Framework Convention on Artificial Intelligence and Human Rights, Democracy and the Rule of Law* (CETS No. 225). Available at: <https://www.coe.int/en/web/conventions/full-list?module=treaty-detail&treaty-num=225> (Access 2.09.2025).
- Council (2025) of Europe. Available at: <https://www.coe.int/en/web/portal/> (Access 2.09.2025).
- Europe (2010) 2020. A strategy for smart, sustainable, and inclusive growth. COM (2010) 2020 final. Brussels, 3.03.
- European (2017) Parliament resolution of 16 February 2017 with recommendations to the Commission on Civil Law Rules on Robotics (2015/2103(INL)). OJ C 2018/252/25.
- European (2018) Commission. *Communication from the Commission to the European Parliament, the European Council, the Council, the European Economic and Social Committee and the Committee of the Regions. Artificial Intelligence for Europe.* COM(2018) 237 Brussels, 25.04.
- European (2019) Commission. *Communication from the Commission to the European Parliament. The Council, the Economic and Social Committee and the Committee of the Regions. Building Trust in Human-Centric Artificial Intelligence.* COM(2019) 168 final. Brussels 8.04.
- European (2020a) Commission. *White Paper On Artificial Intelligence – A European approach to excellence and trust.* COM(2020) 65 final. Brussels 19.02.

- European (2020b) Commission. *Communication from the Commission to the European Parliament, the Council, the Economic and Social Committee and the Committee of the Regions. A European strategy for data.* COM(2020) 66 final. Brussels. 19.02.
- European (2020c) Commission. *Communication from the Commission. Europe 2020. A strategy for smart, sustainable and inclusive growth.* COM(2010) 2020 final. Brussels 3.03.
- European (2020d) Commission. *Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions. A European strategy for data.* COM(2020) 66 final. Brussels. 19.02.
- European (2024a) Commission. *Shaping Europe's digital future. High-level-expert-group on artificial intelligence.* Available at: <https://digital-strategy.ec.europa.eu/en/policies/expert-group-ai> (Access 2.09.2025).
- European (2024b) Commission. *Shaping Europe's digital future. The European AI Alliance.* Available at: <https://digital-strategy.ec.europa.eu/en/policies/european-ai-alliance> (Access 2.09.2025).
- European (2025a) Court of Human Rights. Available at: <https://www.echr.coe.int/> (Access 2.09.2025).
- European (2025b) Commission. *Shaping Europe's digital future. General-Purpose AI Code of Practice.* Available at: <https://digital-strategy.ec.europa.eu/en/policies/ai-code-practice> (Access 2.09.2025).
- European (2025c) Commission. *CORDIS – EU research results. Horizon 2020. A European AI On Demand Platform and Ecosystem.* Available at: <https://cordis.europa.eu/project/id/825619> (Access 2.09.2025).
- Flasiński, M. (2016) *Introduction to Artificial Intelligence.* Springer International, Switzerland. Available at: https://books.google.pl/books?id=UpvvDAAAQBAJ&printsec=frontcover&hl=pl&source=gbs_ge_summary_r&cad=0#v=onepage&q&f=false (Access 2.09.2025).
- Hodges, A. (2014) *Alan Turing: Enigma* (translation from English by W. Bartol). Warsaw: Albatros Andrzej Kuryłowicz S.C., 715 pp.
- ITU (International Telecommunication Union) (2025) *AI for Good.* Available at: <https://aiforgood.itu.int/> (Access 2.09.2025).
- Mishra, V. (2024) *The UN General Assembly on Thursday adopted a landmark resolution on the promotion of “safe, secure and trustworthy” artificial intelligence (AI) systems that will also benefit sustainable development for all.* 21.03. Available at: <https://news.un.org/en/story/2024/03/1147831> (Access 2.09.2025).
- Oppy, G. (2021) *The Turing test in Stanford Encyclopedia of Philosophy Archive.* Available at <https://plato.stanford.edu/archIves/fall2021/entries/turing-test/> (Access 2.09.2025).

- Regulation (2021) (EU) 2021/241 of the European Parliament and of the Council of 12 February 2021 establishing the Recovery and Resilience Facility. OJ 2021 L 57.
- Regulation (2022a) (EU) 2022/1925 of the European Parliament and of the Council of 14 September 2022 on contestable and fair markets in the digital sector and amending Directives (EU) 2019/1937 and (EU) 2020/1828 (Digital Markets Act) (Text with EEA relevance). OJ 2022 L 265.
- Regulation (2022b) (EU) 2022/2065 of the European Parliament and of the Council of 19 October 2022 on a Single Market For Digital Services and amending Directive 2000/31/EC (Digital Services Act) (Text with EEA relevance). OJ 2022 L 27.
- Regulation (2023a) (EU) 2023/1781 of the European Parliament, and of the Council of 13 September 2023 establishing a framework of measures for strengthening Europe's semiconductor ecosystem and amending Regulation (EU) 2021/694 (Chips Act) (Text with EEA relevance). OJ 2023 L 229.
- Regulation (2023b) (EU) 2023/2854 of the European Parliament and of the Council of 13 December 2023 on harmonised rules on fair access to and use of data and amending Regulation (EU) 2017/2394 and Directive (EU) 2020/1828 (Data Act) (Text with EEA relevance). OJ 2023 L 2854.
- Regulation (2024a) (EU) 2024/1183 of the European Parliament and the European Council of 11 April 2024 amending Regulation (EU) No 910/2014 as regards establishing the European Digital Identity Framework. OJ 2024 L 1183.
- Regulation (2024b) (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonised rules on artificial intelligence and amending Regulations (EC) No 300/2008, (EU) No 167/2013, (EU) No 168/2013, (EU) 2018/858, (EU) 2018/1139 and (EU) 2019/2144 and Directives 2014/90/EU, (EU) 2016/797 and (EU) 2020/1828 (Artificial Intelligence Act) (Text with EEA relevance). OJ 2024 L 1689.
- Schneppat, J.O. (2025) *John McCarthy in AI VIPs. Pioneers in the field of AI*. Available at: <http://aivips.org/john-mccarthy/> (Access 2.09.2025).
- Słownik PWN (2025) Słownik Języka Polskiego PWN: *sztuczna inteligencja*. Available at: <https://sjp.pwn.pl/szukaj/sztuczna%20inteligencja.html> (Access 2.09.2025).
- United Nations (2015a) Digital Library. *Transforming our world: the 2030 Agenda for Sustainable Development: resolution/adopted by the General Assembly*. Available at: <https://digitallibrary.un.org/record/3923923?v=pdf> (Access 2.09.2025).

- United Nations (2015b) *Transforming our world: The 2030 agenda for sustainable development* (A/Res/70/1). Available at: <https://sustainabledevelopment.un.org/content/documents/21252030%20Agenda%20for%20Sustainable%20Development%20web.pdf?ref> (Access 2.09.2025).
- United Nations (2021) Digital Library. *Impact of rapid technological change on the achievement of the Sustainable Development Goals and targets: resolution/adopted by the General Assembly*. Available at: <https://digitallibrary.un.org/record/3937141?v=pdf> (Access 2.09.2025).
- United Nations (2023a) Digital Library. *Information and communications technologies for sustainable development: resolution/adopted by the General Assembly*. Available at: <https://digitallibrary.un.org/record/4031524?v=pdf> (Access 2.09.2025).
- United Nations (2023b) Digital Library. *Science, technology and innovation for sustainable development: resolution/adopted by the General Assembly*. Available at: <https://digitallibrary.un.org/record/4031991?v=pdf> (Access 2.09.2025).
- United Nations (2024a) Digital Library. *General Assembly. Seizing the opportunities of safe, secure and trustworthy artificial intelligence systems for sustainable development* (A/78/L.49). Available at: <https://digitallibrary.un.org/record/4043244/?v=pdf> (Access 2.09.2025).
- United Nations (2024b) *General Assembly adopts landmark resolution on artificial intelligence*. Available at: <https://news.un.org/en/story/2024/03/1147831#:~:text=The%20UN%20General%20Assembly%20on%20Thursday%20adopted%20a,that%20will%20also%20benefit%20sustainable%20development%20for%20all> (Access 2.09.2025).
- United Nations (2024c) Office for Digital and Emerging Technologies. *Global Digital Compact*. Available at: <https://www.un.org/digital-emerging-technologies/global-digital-compact#:~:text=The%20Global%20Digital%20Compact%20is%20a%20comprehensive%20framework,global%20governance%20of%20digital%20technology%20and%20artificial%20intelligence> (Access 2.09.2025).
- United Nations (2025a) Department of Economic and Social Affairs. *Sustainable Development. 17 Goals*. Available at: <https://sdgs.un.org/goals> (Access 2.09.2025).
- United Nations (2025b) *Universal Declaration of Human Rights*, 10 December 1948. Available at: <https://www.un.org/en/about-us/universal-declaration-of-human-rights> (Access 2.09.2025).
- United Nations (2025c) Office for Digital and Emerging Technologies: *High-Level Advisory Body on Artificial Intelligence*. Available at: <https://www.un.org/digital-emerging-technologies/ai-advisory-body> (Access 2.09.2025).

- Veisdal, J. (2019) *The Birthplace of AI. An essay about the 1956 „Darmouth workshop”*. Available at: <https://www.cantorsparadise.org/the-birthplace-of-ai-9ab7d4e5fb00/> (Access 2.09.2025).
- von der Leyen, L. (2019) *Speech in the European Parliament Plenary Session. Strasbourg 27.11.* Available at: <https://www.statewatch.org/media/documents/news/2019/nov/eu-com-von-der-leyen-ep-vote-speech-27-11-19.pdf> (Access 2.09.2025).