RAIS RESEARCH ASSOCIATION for INTERDISCIPLINARY STUDIES
November 21-22, 2024

DOI:10.5281/zenodo.14514508

Artificial Intelligence and the Need for Standards and Accountability for Protecting Human Rights, Democracy and the Rule of Law

Titus Corlăţean

Associate Professor PhD, "Carol I" National Defense University of Bucharest, Romania
President of the Foreign Policy Committee of the Romanian Senate, Romania

ABSTRACT: Within the last few years, there has been a diversity of information on the revolutionary industry of Artificial Intelligence: daily news, online articles, statements from investors and policy makers, reports and guidelines published by international organizations, new books and academic works. All of these have occurred with impressive dynamics, involving more and more people. This rich amount of data and information has brought about historical opportunities, significant progress in different activities, as well as developments in national policies. On the other hand, it has increased awareness of potential risks and raised the accountability of both public and private actors in providing a clear and effective regulation to protect human rights, democracy, and the rule of law within this innovative domain. Many countries have already taken legislative measures and issued rules on Artificial Intelligence. This progress marks a promising start for a global approach since the potential risks are not limited to a country or a region – the speed and inherent nature of AI can lead to a rapid spread of its negative effects worldwide, with irreversible consequences for humanity. To prevent these outcomes, significant efforts have been made at the international level to define, assess and address the risks as well as to enforce rules for governing risk management, while continuing to promote AI benefits and development. The present study aims to acknowledge the key role of the Council of Europe Framework Convention on Artificial Intelligence and Human Rights, Democracy and the Rule of Law as the first-ever international treaty on Artificial Intelligence in setting the standards for safely developing and using the AI for the benefit of humanity. It also highlights the important contribution of each party involved to what is now a coordinated path toward a safer use of Artificial Intelligence and trustworthy AI systems.

KEYWORDS: artificial intelligence, human rights, rule of law, multilateralism, Framework Convention of the Council of Europe, opportunities, risks

General Considerations

Artificial Intelligence (AI) has become a reality and is developing with incredible rapidity. It is also extending its applicability not only in different industrial branches but also in other spheres of human activities and human existence - it has become an important tool from the economic, political, social, and health perspectives.

AI continues to develop, thus having an increasing impact on everyone's daily life while paving the way for unknown opportunities of improving people's living conditions. It can also help improve the natural and social environment due to its power to process huge amounts of data, which can support better understanding and better decisions.

On the other hand, it has become obvious that there is a stringent need for regulation within this field, taking into account its vast applicability, possible undesirable interventions within the decisions making process and within the individual autonomy and privacy, as well as all the possible challenges for the future of our society.

How to preserve the benefits and mitigate the risks for humanity? This issue has been at the core of the activity of the Council of Europe, as it is the leading organization dedicated to protecting human rights, democracy, and the rule of law. As a result, The Committee of Ministers of the Council of Europe has adopted, during its 133rd Session of the Committee of Ministers, held in Strasbourg on May 17, 2024, *The Council of Europe Framework Convention on Artificial Intelligence and Human Rights, Democracy and the Rule of Law* (CM(2024)52-

addfinal) **being the first ever international treaty on Artificial Intelligence.** Consequently, this important document deserves special attention, and this is exactly what the present work aims to do.

The decision to adopt the first legally binding instrument on AI has been taken after an intensive process of consultations, research, analysis, and cooperation with other international organizations and entities. It has been a historical moment, a major step forward towards a real protection of human rights, democracy, and the rule of law within the context of the development of AI Systems (AISs) while also taking into account the necessity of preserving the benefits of a modern life, based on scientific and technical development. Therefore, the present paper aims to sum up the long way and the progress achieved by some international organizations, also involving different political, social, and private actors, in their efforts to coordinate an integrated process of regulation on Artificial Intelligence.

Starting with several comments on the new *Framework Convention* of the Council of Europe, this evaluation also points out some important steps and different approaches considered before the final form to be drafted, as well as what is envisaged to be achieved in the future.

Finally. yet importantly, this work is addressed to the public in an attempt to overcome the prejudices, fears, and even apocalyptic assessments on AI, and to underline how important is this new globally legal instrument for preserving our common democratic values, which are worthwhile assets hardly earned over time to our benefit.

Defining AI in a multilateral framework

The first definition of AI formally used in a multilateral framework, being the result of international cooperation, has been set up by the Organisation for Economic Co-operation and Development (OECD), within its updated form, issued on 23 November 2023 (Russell, Perset, and Grobelnik 2023), as follows: "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 definition is commonly used within *The Council of Europe Framework Convention on Artificial Intelligence and Human Rights, Democracy and the Rule of Law*, as well as within the *EU Act on Artificial Intelligence* – a new EU Regulation adopted this year.

Similar definitions of AI have been used within national regulations. In the USA, for example, in line with the OECD Recommendation on AI (2019) and ISO/IEC 22989:2022, the **Artificial Intelligence Risk Management Framework** (NIST 2023) also refers to AI systems as: "engineered or machine-based systems that can, for a given set of objectives, generate outputs such as predictions, recommendations, or decisions influencing real or virtual environments. AI systems are designed to operate with varying levels of autonomy".

In other words, AI has been described by Prof Nigel Shadbolt and Roger Hampson, in their recently published book "As If Human – Ethics and Artificial Intelligence", as being: "the branch of computer science that deals with the creation of systems capable of performing tasks that would ordinarily require human intelligence – systems that can reason". The authors also mentioned **two main categories** of Artificial Intelligence Systems, defined in terms of the scientific advancement and technical performances developed until now. First category includes **rules-based systems,** capable of making decisions or suggestions based on specific input and predefined instructions to be followed. The second one refers to **machine learning systems,** able to identify, analyze and discern patterns by processing a significant amount of data. The latter category conducted to the **generative AI,** such as the pre-trained transformers (GPTs) and Large Multi-Modal Models or Large Language Models (LLMs), also named AI chatbots (or the popular product released on the public space in 2022, named ChatGPT).

It is suggestive how these technological leaps are often described, namely by exemplifying the results of AI in chess games: the first defeat of the best chess player at that time by an AI system, and some years later, the contest between two chess playing machines finalized with the big victory of *machine learning algorithm*.

Related stories and reflections on AI

A simple search on internet reveals a large diversity of information on this huge revolutionary industry of Artificial Intelligence. Each day there are fresh news, online articles, statements of investors and policy makers, new reports and guidelines published by the international organizations, more books and academic works, all of these happening with an impressive dynamic and attracting more and more people.

For instance, many newspapers have recently revealed the allegation - directed to Open AI, one of the top companies in this industry - that its new chatbot has artificially reproduced the voice of a renowned actress (Scarlett Johansson). OpenAI previously publicly declared that the company was to present a tool that can accurately copy/reproduce human voices using AI. Despite denying these allegations, Open AI stated, following the disclosure within the online media, that they would withdraw this tool. Open AI was also confronted with other public allegations coming from a media company that used a huge amount of their articles to train its AI algorithm behind ChatGPT.

These transformative tools come with both advantages and disadvantages. Generative AI is able to determine major changes in economy and society, covering different areas, including education, labor market, and healthcare, offering huge opportunities on the one hand, but raising a lot of questions with regard to risks and ethics.

Risks management and the need for a global legally binding instrument

The use of AI means accountability of both parties - the producer companies and the people who benefit from these tools. Thus, it is necessary to raise awareness of the consequences of such technologies and mitigate the risks, whilst further encouraging research and ethical development. For that, an enormous responsibility lies with the governments and the policy makers - to legislate, monitor, and take administrative and strategic measures. In this regard, clear norms should be enacted, including obligations and sanctions.

Besides the transformations produced in industry, medicine, finance, education and other areas, to some extent, there are serious concerns that AI, as it becomes more and more powerful and influential, may disturb the functioning of the democratic institutions and the development of a healthy democratic society. It has become a well-known fact that there are serious risks of malicious attacks or data breaches, as well as interference with the electoral process, and the spreading of misinformation and disinformation in order to influence the electoral processes.

The online platform *Journal of Democracy*, including publications and essays on actual topics, has drawn attention in an article entitled *How AI threatens democracy* that: “just a month after its introduction, ChatGPT, the generative artificial intelligence (AI) chatbot, hit 100 million monthly users, making it the fastest-growing application in history. For context, it took the video-streaming service Netflix, now a household name, three-and-a-half years to reach one million monthly users. But unlike Netflix, the meteoric rise of ChatGPT and its potential for good or ill sparked considerable debate” (Kreps and Kriner 2023). Further, real questions are also asked - if AI generative models “could be misused by students for studies and research”, “could hijack democracy” due to mass inputs that can influence democratic representation, or even if they are to move forward towards “threatening human existence”.

With regard to the interference with the electoral process, it is worth mentioning the example used in the same article by quoting from the US Senate Intelligence Committee Report 2018: ”masquerading as Americans, these (Russian) operatives used targeted advertisements,

intentionally falsified news articles, self-generated content, and social media platform tools to interact with and attempt to deceive tens of millions of social media users in the United States". It was also stated in the US Senate report that these actions, occurred during the electoral campaign at that time, "sought to polarize Americans on the basis of societal, ideological, and racial differences".

Many countries have already taken legislative measures and issued rules on Artificial Intelligence, a good start for a global approach (OECD Artificial Intelligence Papers 2023a). In the United States, besides the guidance offered to the public and private sector by the AI Risk Management Framework, the **Executive Order on AI** was adopted in 2023, followed by new measures of implementation the US Department of Commerce (The White House 2024).

Details on the purpose, stage and evolutions in adopting national strategies on AI are included in the OECD AI Paper, No 3 (OECD 2023): *The State of Implementation of the OECD AI Principles Four Years On*. In short, it is mentioned that: "In 2019, only a few countries had national AI strategies. Canada, Finland, and Japan were among the first to develop national AI strategies, setting targets and allocating budgets in 2017. Australia, Denmark, France, Germany, Korea, and the United States followed suit in 2018 and 2019. In 2020 and 2021, additional countries announced national AI strategies, including Brazil, Chile, Spain, the Republic of Türkiye, the United Kingdom, and Ukraine. In 2022, Belgium, Israel, Italy and Thailand launched their national AI strategies, while Croatia, Greece, Iceland, and Romania are establishing theirs. Some countries, such as Canada, France and Germany, have updated their national AI strategies, taking stock of achievements and keeping pace with technical, societal and economic developments. As of May 2023, 51 countries had reported a national AI Strategy to the database of national AI policies in OECD.AI from all regions in the world."

It is also important to add what the political debates at the multilateral level have concluded, namely that different standards in different countries may conduct to bypassing ethics and favour the development of artificial intelligence in countries with more favourable approach and lower ethical standards. Moreover, the potential risks are not limited to a country or a region. The dynamic and nature of AI could determine a rapid spread of its negative effects worldwide and might even affect human civilization.

To prevent these from happening, extraordinary work has already been done at international level, for setting up rules and governing risk management, trying to define, assess and address the risks, while continuing to promote AI benefits and development. Each party involved has made its contribution to what is now a coordinated path toward safer use of Artificial Intelligence and trustworthy AI systems.

If someone were interested in learning what has been achieved at the multilateral level for assessing the risks, impact and how to better regulate AI, a good start would be to dig into the OECD's work. This significant volume of documents, analyses and statistics has been possible through close cooperation with different platforms, institutions, independent experts and other representatives. Thus, the OECD has developed the first guidance for a safe use of the AI, especially with regard to properly ensuring the risks management at each stage of the AI system lifecycle. In this respect, it's worth mentioning the important activity of research, analysis and coordination of the **OECD's Expert Group on AI Risk & Accountability** assessing the risks, developing AI core standards, and identifying tools and mechanisms of implementation and control.

This network of experts gathers representatives of the International Organization for Standardization (ISO), the Institute of Electrical and Electronics Engineers (IEEE), the National Institute of Standards and Technology (NIST), the European Committee for Electrotechnical Standardization (CEN-CENELEC), the European Commission (EC), Council of Europe (CoE), UNESCO, OECD, EU-US Trade and Technology Council (TTC) and Responsible AI Institute (RAII)-WEF and other important actors at national and international level.

All these aspects have opened the perspective of a **legally binding instrument** aimed to ensure that AI upholds common standards in human rights, democracy and the rule of law, and minimize the existing risks.

Council of Europe's perspective on artificial intelligence. The role of the Parliamentary Assembly (PACE)

When it comes to protecting human rights, democracy, and the rule of law, the Council of Europe is unanimously recognized, due to its foundation, history and achievements, as the key international organization. Based on its bodies, experts and complex mechanisms, it has been able to set up the appropriate standards on this matter, to update them and to pioneer new ones in line with the evolution of society. It also successfully supports countries in implementing these standards, as well as in preserving them by monitoring the respect of the engagements undertaken by national authorities.

For this reason, the **Council of Europe has taken the lead in a broad coordination process and issued the first treaty on Artificial Intelligence** - *The Framework Convention on Artificial Intelligence and Human Rights, Democracy and the Rule of Law (CETS No. 225)*. Since it has been thought of as a globally applicable legal document, *the Framework Convention (FC)* is also open for accession by non-member States. As provided by the CM(2024)52-final Decision, *the Framework Convention* has been opened for signature starting from 5 September 2024, when the informal Conference of Ministers of Justice for member states took place in Vilnius. On this occasion, Andorra, Georgia, Iceland, Norway, the Republic of Moldova, San Marino, the United Kingdom, Israel, the United States of America and the European Union have signed the Framework Convention.

The text of the Convention has been drafted within the Committee on Artificial Intelligence (CAI) of the Council of Europe. During this process, other international organizations (OECD, OSCE, UNESCO) as well as experts of the European Union, representatives of several countries that are not members of the Council of Europe, representatives of NGOs, private companies, and academic institutions, all had the opportunity to come up with their position and proposals on the text.

By adopting the Framework Convention, the decision-makers did not intend to issue a unique document - it should not be perceived as *the only one treaty*. It sets forth principles and general provisions. Therefore, other binding or non-binding instruments concerning the use of AI in specific sectors or certain standards are further needed. As a result of negotiations, and considering different political approaches and different legal systems, the draft text allows a margin of flexibility in implementation, requesting governments to take measures, graduated and differentiated in terms of needs.

Moreover, in order to ensure an effective implementation of its provisions, the Framework Convention sets up a **follow-up mechanism** and is laying the foundation for further **co-operation**. Thus, it provides the possibility of periodic consultations within the framework of the *Conference of Parties*, in any case when it is convened by a majority of Parties or at the request of the Secretary General of the Council of Europe.

With regard to the next steps to be followed, they will focus on "the **methodology for the risk and impact assessment of AI systems on human rights, democracy and rule of law,** which will be another milestone instrument complementing and operationalizing the Framework Convention and other legal instruments" (Statement of Thomas Schneider 2024).

In line with the statutory role of this Organisation - to safeguard the common ideals and principles and facilitating economic and social progress, and ensure a stronger unity of the member states - the governments of the Council of Europe's member states agreed on a set of **principles** as well as the respective obligations that should be considered in drawing the future policies on AI, as follows:

- To respect human dignity and individual autonomy
- To ensure that the transparency and oversight requirements are observed and are adapted to the specific contexts and risks
- To ensure accountability and responsibility for adverse impacts on human rights, democracy and the rule of law
- To respect equality, including gender equality, and the prohibition of discrimination
- To ensure that privacy rights of individuals and their personal data are protected.

According to the Explanatory Report to the FC adopted by the Committee of Ministers, CM(2024)52-addfinal: "**no provision** of this Framework Convention **is intended to create new human rights** or human rights obligations or undermine the scope and content of the existing applicable protections". It aims at "setting out various legally binding obligations, to facilitate the effective implementation of the applicable human rights obligations of each Party in the context of the new challenges raised by artificial intelligence". The details of what these principles mean in relation to artificial intelligence are presented in the Explanatory Report of the Committee of Ministers. By setting up the principles and obligations enshrined in the text, some important aspects have been considered, aimed at promoting, on a long term, trustful AISs, which are also meant to be broadly accepted by the public. It is to mention, for example, the dynamic and omnipresent character of the AISs, thus being set forth the obligation of "ensuring adequate quality and security *throughout the lifecycle of artificial intelligence systems*"[1].

This provision is essential as different analyses have shown that there could be even more potential harm caused by the AISystems during their use (failures, side-effects, failures, misuse, security breaches) than by design. Therefore, the FC covers, "not only current, but future risks." (Council of Europe Treaty Series CM(2024).

With regard to the **risks**, even if these are not clearly defined within the *Framework Convention*, this treaty provides the obligation for each State Party to "adopt or maintain measures for the identification, assessment, prevention and mitigation of risks posed by artificial intelligence systems by considering actual and potential impacts to human rights, democracy and the rule of law".

It is also to be mentioned that, according to the Council of Europe's Statute, the matter of **national security and defence** is not within the sphere of activity of the organization. Therefore, the *Framework Convention* sets forth that its provisions should not be required to apply to activities "related to the protection of its national security interests", and that "matters related to the national defence do not fall within the scope of the Convention."

The Framework Convention also provides that it shall not apply "to research and development activities regarding AISs not yet available for use."

PACE Opinion on the Draft Framework Convention on Artificial Intelligence and Human Rights, Democracy and the Rule of Law

At the request of the Committee of Ministers, The Parliamentary Assembly of the Council of Europe has issued ***Opinion 303 (2024)*** *on the Draft Framework Convention on Artificial Intelligence and Human Rights, Democracy and the Rule of Law,* adopted, after a debate under urgent procedure, during the second part-session held in Strasbourg on 15-19 April 2024.

It must be pointed out that, based on its own experience and considering some critical positions and other proposals expressed by different stakeholders, the Parliamentary Assembly submitted to the attention of the Committee of Ministers several amendments on the Draft text of the Convention. The most important of these amendments referred to the domains of

[1] OCDE defines *AI system lifecycle* as including the following iterative steps: *design, data and models* (a context-dependent sequence encompassing planning and design, data collection and processing, as well as model building); *verification and validation; deployment; operation and monitoring*. The decision to retire an AI system from operation may occur at any point during the operation and monitoring phase.

national defence, as well as **research and development**, with the proposal to be covered by the *Framework Convention*. Unfortunately, due to some other considerations, this proposal has not been included in the final text. However, further dialogue would be expected to address these important issues, based on other future instruments.

Moreover, the Parliamentary Assembly concluded in its *Opinion* that the text of the Convention **"does not cover to an equal extent public and private actors"[2] and considered this a regrettable fact**. Therefore, PACE calls on the Parties, when ratifying the Framework Convention, to recognize the full applicability of its principles and obligations to activities of private actors. The FC leaves open this possibility, based on a Declaration to be signed by the Party at the time of accession/ratification.

On the other hand, in its Opinion, the Assembly considered with satisfaction that most of the key ethical principles endorsed in its 2020 reports were reflected in the text of the Convention, though some of the principles (for instance, privacy, equality) "are seen by the PACE members as positive individual rights rather than general principles".

It should also be noted that prior to the adoption of the Framework Convention by the Council of Europe, the Parliamentary Assembly has focused its work, besides other topics of interest, on Artificial Intelligence and its impact on human rights, democracy and the rule of law. Thus, in 2020, PACE members adopted within the plenary session several reports focusing on the impact and possible threats of AI to democratic governance, fundamental freedoms, policing justice systems, labor markets, and healthcare. Two other reports referred to the necessity of preventing discrimination caused by AI and the legal aspects of 'autonomous vehicles.

For instance, in the REC 2181 (2020) on the Need for democratic governance of artificial intelligence, the PACE members expressed their support for a Convention of the Council of Europe on artificial intelligence, and recommended to the Committee of Ministers to "**ensure that such a legally binding instrument is based on a comprehensive approach, deals with the whole life cycle of AI-based systems, is addressed to all stakeholders and includes mechanisms to ensure the implementation of this instrument".**

All the recommendations included in these reports referred to the Committee of Ministers, have constituted an important basis and a starting point for the intensive process of negotiation and drafting of the Framework Convention.

The path toward a globally applicable legal framework

As mentioned before, *The Council of Europe Framework Convention on Artificial Intelligence and Human Rights, Democracy and the Rule of Law* is a result of an integrated/coordinated approach to AI among different international and supranational Organizations and fora, including The Organisation for Economic Co-operation and Development and the European Union. Therefore, the contribution of other organizations, entities, and fora must be recalled in this regard.

OECD perspective

The OECD AI Principles are the first intergovernmental standard on Artificial Intelligence, which was adopted in 2019 (OECD 2024). They include concrete recommendations for public policy and strategy, and their general scope ensures they can be applied to AI developments around the world. One of the values-based principles of the OECD AI principles is **Human rights and democratic values, including fairness and privacy.**

[2] According to the article 2, para 1, the Framework Convention applies particularly to the activities undertaken by public authorities, or private actors acting on their behalf. With regard to the **private actors**, FC provides that each state party should submit a Declaration including mentions on how the members states will formally engage to fulfil its obligation and apply the provisions of the Convention to activities of private actors.

AI actors should respect the rule of law, human rights, democratic and human-centered values throughout the AI system lifecycle. These include non-discrimination and equality, freedom, dignity, autonomy of individuals, privacy and data protection, diversity, fairness, social justice, and internationally recognized labor rights. This also includes addressing misinformation and disinformation amplified by AI while respecting freedom of expression and other rights and freedoms protected by applicable international law.

To this end, AI actors should implement **mechanisms and safeguards**, such as capacity for human agency and oversight, including addressing risks arising from uses outside of intended purpose, intentional misuse, or unintentional misuse in a manner appropriate to the context and consistent with the state of the art.

AI should be developed in accordance with human-centered values, such as fundamental freedoms, equality, and fairness, rule of law, social justice, data protection and privacy, as well as consumer rights and commercial fairness.

Some applications or uses of AI systems have implications for human rights, including risks that human rights (as defined in the Universal Declaration of Human Rights) and human-centered values might be deliberately or accidently infringed. It is therefore important to promote "values-alignment" in AI systems (i.e., their design with appropriate safeguards) including capacity for human intervention and oversight, as appropriate to the context. This alignment can help ensure that AI systems' behaviors protect and promote human rights and align with human-centered values throughout their operation. Remaining true to shared democratic values will help strengthen public trust in AI and support the use of AI to protect human rights and reduce discrimination or other unfair and/or unequal outcomes.

These principles also acknowledge the role of measures such as human rights impact assessments (HRIAs) and human rights due diligence, human determination (i.e., a "human in the loop"), codes of ethical conduct, or quality labels and certifications intended to promote human-centered values and fairness.

The OECD AI principles were updated in May 2024 as a response to recent developments in terms of AI technologies. The key elements of the revision, which ensure that the Principles remain relevant, robust, and fit for purpose, are:

- Addressing safety concerns, so that if AI systems risk causing undue harm or exhibit undesired behavior, robust mechanisms and safeguards exist to override, repair, and/or decommission them safely;
- Reflecting the growing importance of addressing mis- and disinformation, and safeguarding information integrity in the context of generative AI;
- Emphasizing responsible business conduct throughout the AI system lifecycle, involving cooperation with suppliers of AI knowledge and AI resources, AI system users, and other stakeholders;
- Clarifying the information regarding AI systems that constitute transparency and responsible disclosure;
- Explicitly referencing environmental sustainability, a concern that has grown considerably in importance over the past five years;
- Underscoring the need for jurisdictions to work together to promote interoperable governance and policy environments for AI, as the number of AI policy initiatives worldwide surges.

The OECD Framework for Classifying AI Systems was developed by AI Network of Experts as a tool for policy-makers, regulators, legislators and others so that they can assess the opportunities and risks that different types of AI systems present and to adapt their national AI strategies.

The OECD Due Diligence Guidance for Responsible Business Conduct (adopted in 2018) provides practical support to enterprises on the implementation of the OECD Guidelines for Multinational Enterprises by providing plain-language explanations of its due diligence

recommendations and associated provisions. Implementing these recommendations can help enterprises avoid and address adverse impacts related to workers, human rights, the environment, bribery, consumers and corporate governance that may be associated with their operations, supply chains and other business relationships. The Guidance includes additional explanations, tips and illustrative examples of due diligence.

This Guidance also seeks to promote a common understanding among governments and stakeholders on due diligence for responsible business conduct. The UN Guiding Principles on Business and Human Rights, as well as the ILO Tripartite Declaration of Principles Concerning Multinational Enterprises and Social Policy, also contain due diligence recommendations, and this Guidance can help enterprises implement them.

The EU Artificial Intelligence Act

The EU AI Act has already reached its final stage. The draft Regulation has been unanimously supported by all member states, in December 2023. It has been approved within the Internal Market and Civil Liberties Committees of the European Parliament, and afterward within plenary sitting of the European Parliament, on 13 March 2024, with 523 votes in favor, 46 against and 49 abstentions (European Parliament 2019-2024). Further proceedings have included the so-called *corrigendum procedure*, in line with the Rules of Procedure of the European Parliament (2024a), aimed at identifying and correcting the text, as well as the formal adoption by the Council, occurred on 21 May 2024 (European Council of the EU 2024). Then, the legislative act was published in the EU's Official Journal on 12 July 2024 and entered into force twenty days after this publication, on 2 August.

The EU AI Act will be fully applicable 24 months after the date of its entry into force - from 2 August 2026. Specific dates of implementation are provided for the following situations: bans on prohibited practices; codes of practice; general-purpose AI rules including governance and obligations for high-risk systems (6-36 months after the entry into force date, differing for each category) (European Parliament 2024b).

The EU AI Act sets up provisions with regard to: classifying the risks, defining prohibited AI systems, enforcing rules for High Risks AI systems, requirements for providers of High Risks AI systems and of General Purpose AI Models, rules on market monitoring, market surveillance, governance and enforcement, etc. It covers over 400 pages. Its purpose, according to the text, is to "improve the functioning of the internal market by laying down a uniform legal framework in particular for the development, the placing on the market, the putting into service and the use of artificial intelligence systems (AI systems) in the Union, in accordance with Union values". It also aims "to promote the uptake of human centric and trustworthy artificial intelligence (AI) while ensuring a high level of protection of health, safety, fundamental rights as enshrined in the Charter of fundamental rights of the European Union (the 'Charter'), including democracy, the rule of law and environmental protection, against the harmful effects of AI systems in the Union, and to support innovation".

It is also to be mentioned that *the EU AI Act* does not apply to for AISs used for **military and defence** as well as for **research purposes**.

There is much more to say about the EU AI Act, but due to the vast and complex approach of this Regulation, it should be separately analyzed. It needs special attention and a deep study in order to understand all the issues it addresses and how it will be implemented on the market.

From the Resolutions of the UN General Assembly to the UN Global Digital Compact

On 21 March 2024, the UN General Assembly (2024) adopted a **Resolution on safe, secure and trustworthy artificial intelligence systems for sustainable development – RES/78/265**. The document was adopted by consensus, having the support of all 193 UN member nations. The Resolution recognizes that "safe, secure and trustworthy artificial intelligence systems 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**".

It also encourages Member States and invites multi-stakeholders from all regions and countries, within their respective roles and responsibilities, to develop and support regulatory and governance approaches and frameworks related to safe, secure and trustworthy artificial intelligence systems.

Moreover, the Resolution underlines the need to promote an international approach to AI, and to implement "effective, internationally interoperable safeguards, practices, standards and tools for artificial intelligence designers, developers, evaluators, deployers, users and other stakeholders for safe, secure and trustworthy artificial intelligence systems".

Last, but not least, the UNGA Resolution calls for cooperation of all parties involved - Governments, the private sector, civil society, international and regional organizations, academia and research institutions and technical communities and all other stakeholders - to this end.

The **RES/78/265** has been strengthened by another resolution adopted by the UNGA on 1 July 2024, under the title *Enhancing international cooperation on capacity building of AI (RES/78/311),* focusing on **the principle of artificial Intelligence for good for all**. It emphasizes that "the Member States should enjoy equal opportunities in the design, development, deployment, decommissioning and the use of the artificial intelligence". It also encourages Governments to take necessary measures and eliminate discrimination and all barriers faced by those in vulnerable situations.

Moreover, the United Nations adopted in unanimity in September 2024, during the Summit of the Future, the *Global Digital Compact* - a relevant document setting up important recommendations with regard to enhancing international governance of artificial intelligence for the benefit of humanity.

These documents should be considered a very good signal for a large consensus, at a global scale, with regard to the necessity, importance and benefits of enforcing legally binding instruments at the international level.

Conclusions

In dealing with the multiple consequences of the implementation of Artificial Intelligence, international organizations took the initiative to adopt a set of documents and a legal framework aiming to ensure a proper development of these new technologies, in full respect of fundamental rights and to mitigate the potential risks.

The adoption of the *Council of Europe Framework Convention on Artificial Intelligence and Human Rights, Democracy and the Rule of Law* should be considered a major achievement. It is the result of a remarkable process of analysis, research, consultation and cooperation with other international organizations and entities, taking into account the proposals of different stakeholders and experts.

This **first-ever international treaty on Artificial Intelligence,** laying down a set of **principles** and respective obligations that should be considered by the Member States in drawing the future policies on AI, has opened the way for other specific rules and international regulation on this matter**.** Thus, shortly after, the European Union adopted the EU Artificial Intelligence Act. The new European Regulation will ensure clear provisions with regard to classification of AI systems as high-risk and prohibited AI practices; obligations for General-purpose AI models (GPAI) and limited risk AISs; penalties; governance; other important aspects related to the design, use and monitor of the AI Systems in full respect of fundamental rights.

At a global scale, by adopting the Global Digital Compact in September 2024, the Member States of the United Nations have established as a main objective to enhance international governance of artificial intelligence for the benefit of humanity.

In the coming period, special attention should be given to the implementation process of the new rules, as well as to promote at a large scale the respect of these principles and obligations. Another important issue would be to ensure proper information to the public. Further debates should also be focused on how parliamentarians can act in order to promote a safe, human-centered, transparent, clearly regulated and monitored Artificial Intelligence. They can have an important role in connecting national legislative frameworks with regional and global initiatives on Artificial Intelligence, as well as in keeping the people informed with regard to the advantages of rules-based AISs, and how to get information and advice.

The digitalization process has started in many national parliaments and, based on an extended exchange of good practices, other legislative bodies should be supported in using the advantages of the new technologies within the parliamentary work. On the other hand, within the context of the advancement of Generative AI, further debates could be held on how to deal with these AI Systems within the parliamentary work. More debates within the parliaments on AI regulation will provide real benefits and contribute to proper solutions for our society.

References

Committee of Ministers of the Council of Europe. 2024. "The Council of Europe Framework Convention on Artificial Intelligence and Human Rights, Democracy and the Rule of Law." https://rm.coe.int/1680afae3c.

Council of Europe Treaty Series - No. 225. 2024. "Explanatory Report to the Council of Europe Framework Convention on Artificial Intelligence and Human Rights, Democracy and the Rule of Law," adopted by the Committee of Ministers, CM(2024)52. Available at https://rm.coe.int/1680afae67.

European Council of the EU. 2024. "Artificial intelligence (AI) act: Council gives final green light to the first worldwide rules on AI." Available at https://www.consilium.europa.eu/en/press/press-releases/2024/05/21/artificial-intelligence-ai-act-council-gives-final-green-light-to-the-first-worldwide-rules-on-ai/.

European Parliament. 2019-2024. Artificial Intelligence Act. European Parliament legislative resolution of 13 March 2024 on the proposal for a regulation of the European Parliament and of the Council on laying down harmonised rules on Artificial Intelligence (Artificial Intelligence Act) and amending certain Union Legislative Acts (COM(2021)0206 – C9-0146/2021 – 2021/0106(COD)). https://www.europarl.europa.eu/doceo/document/TA-9-2024-0138_EN.pdf.

European Parliament. 2024a. "Rules of Procedure of the European Parliament." Available at https://www.europarl.europa.eu/doceo/document/RULES-9-2023-11-01-RULE-241_EN.html.

European Parliament. 2024b. "Artificial Intelligence Act: MEPs adopt landmark law." Available at https://www.europarl.europa.eu/news/en/press-room/20240308IPR19015/artificial-intelligence-act-meps-adopt-landmark-law.

Kreps, Sarah and Doug Kriner. 2023. "How AI Threatens Democracy." Journal of Democracy. 34(4): 122-131. https://www.journalofdemocracy.org/articles/how-ai-threatens-democracy/.

National Institute of Standards and Commerce of the US Department of Commerce (NIST). 2023. "Artificial Intelligence Risk Management Framework (AI RMF)." Available at https://nvlpubs.nist.gov/nistpubs/ai/NIST.AI.100-1.pdf.

OECD (Organisation for Economic Co-operation and Development). 2024. Recommendation of the Council on Artificial Intelligence, which was the first intergovernmental standard on AI, approved during the Meeting of the Council at Ministerial Level (2-3 May 2024). C/MIN(2024)16/FINAL. Available at https://one.oecd.org/document/C/MIN(2024)16/FINAL/en/pdf.

OECD Artificial Intelligence Papers. 2023. Emerging Trends in AI Skill Demand Across 14 OECD Countries, https://doi.org/10.1787/dee339a8-en. Available at https://www.oecd-ilibrary.org/docserver/7c691b9a-en.pdf?expires=1716740872&id=id&accname=guest&checksum=501200E188BE434C3C572CB0610D00F5.

OECD. 2023. "The state of implementation of the OECD AI Principles four years on." Available at https://www.oecd.org/en/publications/the-state-of-implementation-of-the-oecd-ai-principles-four-years-on_835641c9-en.html.

Russell, Stuart, Karine Perset, Marko Grobelnik. 2023. "Updates to the OECD's definition of an AI system explained." *Oecd.Ai.* Available at https://oecd.ai/en/wonk/ai-system-definition-update.

Statement of Thomas Schneider, Chair of the Committee on Artificial Intelligence of the Council of Europe between 2022-2024. Available at https://www.coe.int/en/web/artificial-intelligence/cai.

The White House. 2024. "Fact Sheet: Key AI Accomplishments in the Year Since the Biden-Harris Administration's Landmark Executive Order." Available at https://www.whitehouse.gov/briefing-room/statements-releases/2024/10/30/fact-sheet-key-ai-accomplishments-in-the-year-since-the-biden-harris-administrations-landmark-executive-order.

UN General Assembly, 2024. "Resolution on safe, secure and trustworthy artificial intelligence systems for sustainable development – RES/78/265." Available at https://documents.un.org/doc/undoc/ltd/n24/065/92/pdf/n2406592.pdf?token=TmSaoCQephXZVYd2dk&fe=true.