WASD
WORLD ASSOCIATION FOR SUSTAINABLE DEVELOPMENT

19

OUTLOOK 2025

ETHICS IN ARTIFICIAL INTELLIGENCE: WHO IS IN CHARGE?

DR PETRU DUMITRIU
DiploFoundation, Switzerland/Malta
Senior Fellow and Lecturer on Multilateral Diplomacy
Email: petrud@diplomacy.edu
ORCID: 0009-0006-2772-7623

ABSTRACT

PURPOSE: This study aims to analyse the perspectives of the United Nations (UN) system, European Union (EU) and other international organisations on the ethical aspects of the use of Artificial Intelligence (AI) tools.

DESIGN/METHODOLOGY/APPROACH: The paper will explore the attempts to define the main ethics challenges that may accompany the use of AI as reflected in the documents adopted by international organisations, in particular by the EU and the UN.

FINDINGS: As the UN does not have its own capacity to develop AI tools, it is playing a considerable role in defining principles, limits, and possible rules on the use of AI and stimulating international co-operation. However, the regulations adopted by the EU on AI may influence international co-operation at the global level.

ORIGINALITY: The paper identifies the most significant actions taken at international level, and analyses their similarities, differences, and neglected areas in dealing with the preservation of a strong ethical dimension and the use of AI in service of the public good.

IMPLICATIONS: The paper will help to develop a better knowledge and understanding of the most meaningful conceptual advances in defining and disseminating the principles of ethics among the many stakeholders involved in developing, promoting, and using AI instruments.

KEYWORDS: *Ethics; AI; Council of Europe Framework Convention; EU AI Act; UNESCO; Unethical Practices; Enforcement; WSIS.*

CITATION: Dumitriu, P. (2025): Ethics in Artificial Intelligence: who is in charge? In Ahmed, A. (Ed.): *United Nations: What Next After 2030 Agenda and SDGs.* World Sustainable Development Outlook 2025, Vol. 21, pp.301-320. WASD: London, United Kingdom.

RECEIVED: 9 October 2025 / **REVISED:** 15 December 2025 / **ACCEPTED:** 16 December 2025 / **PUBLISHED:** 30 December 2025

DOI: 10.47556/B.OUTLOOK2025.23.20

CONTEXT

While ethics in relation to the use of Artificial Intelligence (AI) is the concern of almost everyone, the easiest answer to the question "who is in charge?" is "the industry". To the extent that emerging technologies are new, that is they represent innovation, they are already flying above existing laws, rules, and regulations, which are inherently "old". Innovation in technology is a continuum of spontaneous and unpredictable moves. Each technological advance leads to a new process of dissemination, use and experimentation, which leads to another advance.

The logic of innovation is not substantially altered by ethical, moral, equity, and other societal considerations. Although innovations in technology claim to bring more efficiency, comfort and speed, for the consumers, their ultimate goal is to return the investment, to increase profits, to expand on new or larger markets, etc. Governments and other regulators will inevitably be behind the fait accompli by the technology.

The new technologies will always embed the risk of bad use. The champions of the new technologies will constantly promise and put forward the intention to serve the public good. The risks and possible abuses will be kept in the shade, but they soon materialise in the actual use, when it is more difficult to prevent negative consequences. The process of a social reflection and subsequent norm setting takes time and requires the involvement of many stakeholders, political, social, and economic correlations. It is a race between an elephant and a gazelle.

In many cases related to digital technologies, governments are usually left with one dilemma only: the need for new and specific norms, or the proper adjustment and application of the existing rules. The answers to this dilemma are already given. The first is that the same rights that people have offline must also be protected online. The second is that what is illegal offline should also be illegal online. However, the above conclusions should not be taken in the absolute. Caution is needed and each new technology should be examined on its own merits and perils, even if some conclusions will be repetitive. Starting with the old maxim of Plato: "Good people do not need laws to tell them to act responsibly, while bad people will find a way around the laws". If categorising people as good and bad is difficult enough, doing the same with technology is almost impossible. Industry always pretends that their technology is good. Technology seems to enjoy the presumption of innocence at its birth.

The way we see AI is faced with the same problems and dilemmas. Ethics goes beyond law. The advent of AI is spectacular and has a powerful impact on the way people think about their future. Yet, its progress does not come with a revolution in the way governments alone or in multilateral co-operation settings will react. What it is certain is that the issue is on the global agenda and there is work done or undergoing. Ethics is probably one of the most important aspects to be examined.

DOI: 10.47556/B.OUTLOOK2025.23.20

Drawing the border between ethical and non-ethical use is more difficult than making the difference between legal and illegal.

Once brought to commercial existence, the digital technologies raise multiple safety and security issues. These could have been anticipated, but which the producers and the society at large ignored at the early stage of their emergence. The fact is that technology outstrips the capacity of rapid reaction of the society, at national or international level. When they are faced with the reality, it is often too late for the regulators to envisage preventive safeguards and therefore damage containment is the only option left. Or, worse, there is no consensus on who should write rules and regulations to mitigate those concerns. This problem is painstaking when it is about regulations that should be elaborated at the international level. For example, it took the United Nations system 21 years to draft and adopt a Convention on Cybercrime, although cybersecurity issues are almost by definition global issues (WSIS, 2003; UN, 2024a).

There is no doubt that innovations, AI included, come with a pledge of comfort and convenience which will be brought under the spotlight by their evangelists. Digital technologies have proved to have dark sides, which have shown their ugly faces sooner rather than later. The worldwide web is now the predilect arena to proliferate hate speech, to recruit and create criminal groups, to spread fake news, lies, propaganda.

The drones who were acclaimed as means of providing food to refugee camps or to groups isolated in vulnerable conditions, or to send postal parcels in remoted areas, are now among the main tools of waging wars. They brought indeed more convenience in killing people and destroying properties, with no concern for the number or the faces of the human victims.

Drawing the border between ethical and non-ethical use is more difficult than making the difference between legal and illegal. Before trying to see what has been done in terms of codification of international soft and hard law, we should see first what kind of ethics we could have in mind when applied to the use of AI systems and only after to identify if there is a difference between ethics in general and ethics that should cover the specificities of AI, whichever they are.

DOI: 10.47556/B.OUTLOOK2025.23.20

PART I: ETHICS FOR BEGINNERS: BUSINESS AS USUAL OR AN INFLEXION POINT?

An incursion between various definitions of ethics does not help too much those who try to depict - in one single pronouncement - a clear landscape of those characteristics of AI that may make them prone to break ethical principles. Let us examine a sample of such definitions picked up randomly without pretence of exhaustiveness, and make some ad hoc conclusions on what seems to be relevant for normative purposes.

Moral and morality are also institutional attributions, and so are ethics.

What is ethics for AI?

1. The study of the concepts involved in practical reasoning: good, right, duty, obligation, virtue, freedom, rationality, choice.

This is a classic, academic, all-encompassing definition that appears to work for any kind of human undertaking, its basic being the Greek word ethos, which means, among others, character or characteristics. The term was originally used by Aristotle, as a person's character or personality, in the broader context of rhetoric. In the famous triad, besides logos (logic, reason) and pathos (emotions), ethos would mean credibility and play the role of ensuring a balance between passion and caution.

In contemporary language, ethos has a different meaning and refers to the practices or values that distinguish one person, organisation, or society from others, as manifested in their attitudes and values. Derived from that, ethics is:

2. "A branch of philosophy dealing with what is morally right or wrong".

While the first definition seemed to indicate that ethics referred to an individual, with the second expansion of meaning we moved towards collective relevance (institutions, organisations, communities). Moral and morality are also institutional attributions, and so are ethics.

This assumption is confirmed by the following definition, which extrapolates ethics to both individual and institutions.

3. An initial definition of ethics is the analysis, evaluation, and promotion of correct conduct and/or good character, according to the best available standards. […] Ethics asks what we should do in some circumstances, or what we should do as participants in some form of activity or profession. Ethics is not limited to the acts of a single person. Ethics is

 DOI: 10.47556/B.OUTLOOK2025.23.20

also interested in the correct practices of governments, corporations, professionals, and many other groups (University of Wisconsin-Madison, n.d.).

This description also opens the way to the question of standards. Defining standards is not easy either, but it is done. Various alternatives prove the complexity of claiming the existence of a set of standards covering all requirements that would serve the concept of ethics.

Key International Standardization Organisation (ISO) standards on AI:

1. ISO/IEC 42001:2023 - This is the first international standard specifically focused on AI management systems. It outlines requirements and guidance for establishing, implementing, maintaining, and improving an AI management system within organisations. This standard addresses challenges posed by AI, including ethical considerations, transparency, and continuous learning, thereby helping organisations manage risks and opportunities associated with AI effectively.
2. ISO/IEC 23894:2023 - This standard provides guidance on risk management in AI systems. It emphasises the importance of ensuring that AI algorithms are understandable and can be audited for bias and fairness, which is crucial for building trust in AI applications.

4. An ethic is a framework, or guiding principle, and it's often moral. […] A social ethic might include "treating people as you want to be treated." Used in the plural, ethics refers to the moral rules that you live by.

5. We can think of ethics as the principles that guide our behaviour toward making the best choices that contribute to the common good of all (SCU, n.d.).

6. Ethical theories study human moral behaviour and attempt to discover normative rules or maxims that describe what can be called "right action" and "wrong action." Theories of ethics can be deontological systems, which are built around absolute moral rules that must be followed regardless of the outcome (Erwin, n.d.).

7. Ethics is based on well-founded standards of right and wrong that prescribe what humans ought to do, usually in terms of rights, obligations, benefits to society, fairness, or specific virtues (Velazquez *et al.*, 2010).

8. Ethics is the discipline concerned with what is morally good and bad, right and wrong (Britannica, 1974).

A simple enumeration of the notions that may accompany the ethics - moral and immoral, right and wrong, good and bad, correct and incorrect - would be a clear indication that we are talking about human choices. Consequently, direct attribution of ethical features

DOI: 10.47556/B.OUTLOOK2025.23.20

to any of the present AI systems is overstretched and misleading. Technology as such is supposed to be morally neutral.

Ethical concerns ultimately lie with humans, during all AI design, development, and deployment stages.

The simple notion of "AI ethics" is confusing. The imperative need in using AI systems and applications is the human oversight. Ethics is the responsibility of humans wherever they are. This means that ethical concerns ultimately lie with humans, during all AI design, development and deployment stages. The right approach at this point should culminate with the ethical use of AI.

There are, of course, the dangers and the risks associated with autonomous decision-making by AI, where human oversight will have the higher responsibility to ensure that ethical standards are upheld. History has shown us that humans will never be capable of mastering technology in a way that ensures that it is used only for good and moral purposes. On the contrary, even at this stage an immeasurable amount of talent and creativity is spent in using AI for fake news, manipulation, propaganda and plagiarism.

It is fair to note that international organisations, the United Nations in particular, contextualised the need for ethics whenever they dealt with international co-operation in the use of technologies. One such example is the World Summit on the Information Society and the Geneva Declaration of Principles (UN, 2003) (see text box below).

Ethical Dimensions of the Information Society

The Information Society should respect peace and uphold the fundamental values of freedom, equality, solidarity, tolerance, shared responsibility, and respect for nature.

We acknowledge the importance of ethics for the Information Society, which should foster justice, and the dignity and worth of the human person. The widest possible protection should be accorded to the family and to enable it to play its crucial role in society.

The use of ICTs and content creation should respect human rights and fundamental freedoms of others, including personal privacy, and the right to freedom of thought, conscience, and religion in conformity with relevant international instruments.

All actors in the Information Society should take appropriate actions and preventive measures, as determined by law, against abusive uses of ICTs, such as illegal and other acts motivated by racism, racial discrimination, xenophobia, and related intolerance, hatred, violence, all forms of child abuse, including paedophilia and child pornography, and trafficking in, and exploitation of, human beings.

Source: WSIS, 2003, The Declaration of Principles, section B10

DOI: 10.47556/B.OUTLOOK2025.23.20

In the sections to come, we shall try to identify what realistic measures international organisations undertake to solve problems triggered not by the AI as such, but by the humans who are using and abusing all tools at their possession, including AI.

We do not have to rely on AI to solve our old and new problems: we have to cultivate and strengthen human intelligence, including by using and improving technological tools. This leads us to a conclusion drawn by one of the most careful philosophers of ethics, Spinoza.

> *Without intelligence there is not rational life: and things are only good, in so far as they aid man in his enjoyment of the intellectual life, which is defined by intelligence. Contrariwise, whatsoever things hinder man's perfecting of his reason, and capability to enjoy the rational life, are alone called evil* (Spinoza, 1677).

Now, that we know what is old in ethics, let us see if and what the AI advent brings us in global governance.

PART II: SOFT NORMS OF ETHICS

UNESCO Recommendation on the Ethics of Artificial Intelligence

In November 2021, UNESCO adopted the Recommendation on the Ethics of Artificial Intelligence, marking its first global standard on AI ethics (UNESCO, 2021). This recommendation is applicable to all its 194 member states and emphasises the protection of human rights and dignity as fundamental principles. It advocates for transparency, fairness, and the necessity of human oversight of AI systems. The recommendation outlines policy action areas that could guide policy-makers in translating these core values into actionable frameworks across diverse sectors, including data governance, education, health, and social wellbeing.

One of the merits of UNESCO's recommendation is that it describes the specific challenges brought by AI to ethics in addition to what other technologies may have brought. Indeed, the document enlists a comprehensive list of new types of ethical issues that AI systems raise: impact on decision-making, employment and labour, social interaction, health care, education, media, access to information, digital divide, personal data and consumer protection, environment, democracy, rule of law, security and policing, dual use, and human rights and fundamental freedoms, including freedom of expression, privacy and non-discrimination.

According to UNESCO, less visible threats come from the potential of AI algorithms to reproduce and reinforce existing biases, and thus to exacerbate already existing forms of discrimination, prejudice and stereotyping. The impact on labour is also mentioned as

DOI: 10.47556/B.OUTLOOK2025.23.20

a result of the capacity of AI systems to perform tasks that previously only living beings could do, and that were, in some cases, even limited to human beings only (UNESCO, 2021: Section 1, p.10).

The stages of the AI system lifecycle:

Research, design and development to procurement, deployment and use, including maintenance, operation, trade, financing, monitoring and evaluation, validation, end of use, disassembly and termination.

(see para. 2(b) of UNESCO's Ethics of AI Recommendation)

UNESCO's Recommendation outlines several key principles aimed at ensuring the ethical development and deployment of AI technologies throughout the AI lifecycle. Those key principles could be clustered into two categories.

Impact on Human Rights and Fundamental Freedoms as Defined by International Law

- ***Human rights and dignity:*** The protection of human rights and dignity is fundamental. All AI systems must respect, protect, and promote these rights throughout their lifecycle.
- ***Fairness and non-discrimination:*** AI actors must ensure fairness and non-discrimination in AI systems, actively working to minimise biases and ensure equitable access to AI benefits for all individuals.
- ***Privacy protection:*** Privacy must be safeguarded throughout the lifecycle of AI systems, with robust data protection frameworks established to prevent misuse of personal data.
- ***Necessity and proportionality:*** The use of AI systems should be governed by the principle of necessity and proportionality, ensuring that AI applications are justified and not used for harmful purposes such as social scoring or mass surveillance.

Impact on Other Aspects of Social and Economic Life

- ***Promotion of diversity:*** the importance of respecting and promoting diversity and inclusiveness in all aspects of AI development and application.
- ***Inclusive governance:*** Governance mechanisms for AI must be inclusive, transparent, multidisciplinary, multilateral, and involve diverse stakeholders to ensure comprehensive oversight.

DOI: 10.47556/B.OUTLOOK2025.23.20

- ***Continuous impact assessment:*** There should be ongoing assessments of the human, social, cultural, economic, and environmental impacts of AI technologies to ensure alignment with sustainable development goals.
- ***Public engagement and education:*** Promoting public understanding of AI through education, civic engagement, and training in digital skills.
- ***Transparency and explainability:*** AI systems should be transparent and explainable, allowing users to understand how decisions are made.
- ***Accountability:*** There must be clear accountability mechanisms in place for AI systems, ensuring that ultimate responsibility remains with human actors rather than being displaced by technology.
- ***Safety and security:*** Any threats posed by these systems must be addressed to safeguard human well-being and environmental health.

Notably, UNESCO also launched a Global AI Ethics and Governance Observatory that claims to be "a tool for ethical impact assessment". It is a platform that aims to provide resources for interested stakeholders to navigate amidst the ethical challenges posed by AI technologies. As it is the fruit of the UNESCO recommendation, the Observatory promotes collaboration, knowledge sharing, and capacity building.

More interesting, the Observatory brings to attention case studies "to expose AI systems and tools […] released to users without clear and transparent analysis of the potential risks and how they might be mitigated, even when such risks were foreseeable" (UNESCO, 2023a). Such an example is the study entitled "Foundation models such as ChatGPT through the prism of the UNESCO. Recommendation on the Ethics of Artificial Intelligence" (UNESCO, 2023b). This study signals the danger of those generative AI tools, branded as "experimental" by their developers, that propose large language models that have routinely generated inaccurate, misleading or discriminatory content.

UNESCO's Recommendation prompted us to enter into the work of the United Nations system. Not surprisingly, this does not mean one single door because various entities in the system do not coalesce for a single vision but keep running on different tracks, at the risk of duplication and overlapping.

Principles for the Ethical Use of Artificial Intelligence in the United Nations System

Starting from UNESCO work, the United Nations System Chief Executive Board for Coordination (CEB) produced its own "ethical approach" consisting of ten "Principles for the Ethical Use of Artificial Intelligence in the United Nations system".

DOI: 10.47556/B.OUTLOOK2025.23.20

- Do no harm
- Defined purpose, necessity and proportionality
- Safety and security
- Fairness and non-discrimination
- Sustainability
- Right to privacy data protection and data governance
- Human autonomy and oversight
- Transparency and explainability
- Responsibility and accountability
- Inclusion and participation.

> ***All stages of the AI system lifecycle should follow and incorporate human-centric design practices and leave meaningful opportunity for human decision-making.***

As UNESCO and CEB enouncements are not identical, we would extract from the CEB list three key principles that are different in formulation and content.

Do no harm: AI systems should not be used in ways that cause or exacerbate harm, whether individual or collective, and including harm to social, cultural, economic, natural, and political environments. […] The intended and unintended impact of AI systems, at any stage in their lifecycle, should be monitored in order to avoid causing or contributing to harm.

Sustainability: Any use of AI should aim to promote environmental, economic and social sustainability. To this end, impacts of AI technologies should continuously be assessed and appropriate mitigation and/or prevention measures should be taken to address adverse impacts, including on future generations.

Human autonomy and oversight: The United Nations system organisations should ensure that AI systems do not overrule freedom and autonomy of human beings and should guarantee human oversight. All stages of the AI system lifecycle should follow and incorporate human-centric design practices and leave meaningful opportunity for human decision-making. Human oversight must ensure human capability to oversee the overall activity of the AI system and the ability to decide when and how to use the system in any particular situation, and the ability to override a decision made by a system. As a rule, life and death decisions or other decisions affecting fundamental human rights of individuals must not be ceded to AI systems, as these decisions require human intervention (UNS, 2022).

DOI: 10.47556/B.OUTLOOK2025.23.20

An initiative by the United Nations Secretary-General

The Secretary-General of the United Nations, António Guterres, could not miss the opportunity to claim his own imprint on the list of attempts to define the role of the world organisation in handling AI from a perspective of global governance. The Secretary-General convened a High-Level Advisory Body on Artificial Intelligence whose work has culminated in the adoption of its Final Report: "Governing AI for Humanity"; this report establishes a comprehensive framework for global AI governance.

The report emphasises the need for an inclusive and co-operative approach to AI governance, recognising that current frameworks are insufficient, and that the development of AI is largely controlled by a few multinational companies. The report issued several recommendations for establishing a robust global governance framework, among which:

- Establishing an independent panel to provide reliable scientific knowledge about AI, helping to inform policy decisions.
- Creating a platform to ensure technical interoperability of AI systems across borders, involving various stakeholders including tech companies and civil society.
- Standardising data-related definitions and principles to ensure transparency and accountability in AI systems.

Risks Associated with AI

The report does not deal with the issue of ethics but uses a list of risks associated with the AI which is highly relevant for an ethical perspective.

- Damage to information integrity (mis/disinformation, impersonation)
- Intentional use of AI in armed conflict by state actors (autonomous weapons)
- Inequalities arising from differential control and ownership over AI technologies (increased concentration of wealth/power among individuals, corporations)
- Intentional malicious use of AI by non-state actors (crime, terrorism)
- Discrimination/disenfranchisement, particularly against marginalised communities (use of biased AI in hiring or criminal justice decisions)
- Intentional use of AI by state actors that harms individuals (mass surveillance)
- Human rights violations
- Inaccurate information/analysis provided by AI in critical fields (misdiagnoses by medical AI)
- Intentional use of AI by corporate actors that harms customers/users (hyper-targeted advertising, AI-driven addictive products)

DOI: 10.47556/B.OUTLOOK2025.23.20

- Violation of intellectual property rights
- Environmental harms (accelerating energy consumption and carbon emissions)
- Harms to labour from adoption of AI (disruption of labour markets, increased unemployment)
- Unintended autonomous actions by AI systems (loss of human control over autonomous agents, deceptive/manipulative actions)
- Unintended multi-agent interactions among AI systems (trading AIs engaging in collusive signalling) (UN, 2024b)

The industry will always have an upper hand in all stages of AI life cycles.

AI and Military Uses

The UNESCO Recommendation on the Ethics of Artificial Intelligence appears to be the central piece produced by the United Nations system and recognised as such by the General Assembly in its first resolution on AI, titled "Seizing the opportunities of safe, secure and trustworthy artificial intelligence systems for sustainable development" (UN, 2024c).

However, the adoption of this resolution should not be overestimated. First, we are dealing again only with a recommendation, with a political declaration, with a framework for international co-operation. No binding rules. The industry will always have an upper hand in all stages of AI life cycles. The international machinery will always need more time to keep up with the new developments and will have to beg for resources to do anything meaningful. Just raising the flag of ethics and drawing the border between right and wrong will never be enough.

Second, and even worse, the governments of major powers and their military establishment will have hands free on developing AI systems for military purposes. The resolution 78/265 does not hide this truth and makes it clear: the resolution "refers to artificial intelligence system in the non-military domain" (sixth preamble paragraph). The Pandora's box will be open for ever.

DOI: 10.47556/B.OUTLOOK2025.23.20

PART III: CODIFYING INTERNATIONAL LAW

The Framework Convention on Artificial Intelligence and Human Rights, Democracy, and the Rule of Law

Since 2024, the international community could be satisfied that there is no longer a legal vacuum on AI. The Council of Europe adopted a Framework Convention on Artificial Intelligence and human rights, democracy, and the rule of law (CoE, 2024), the first-ever international legally binding treaty in this field. As the title of the Convention indicates, the Council of Europe places its document in a very clear context and ambit, one that fits its mandate and competence as an organisation: human rights, democracy, and rule of law. Therefore, it is reasonable to expect that, while it is remains technology-neutral, the Convention will not regulate technology but the way the human factors will use it. So, the first reading indicated that the Convention comes as close as it can to the human responsibilities as well as to the meaning of international co-operation on artificial intelligence.

The essential meaning of the convention is candidly explained in two preamble paragraphs; these are a paradoxical juxtaposition that say everything and its opposite:

> *Recognizing that activities within the lifecycle of artificial intelligence systems may offer unprecedented opportunities to protect and promote human rights, democracy and the rule of law;*
>
> *Concerned that certain activities within the lifecycle of artificial intelligence systems may undermine human dignity and individual autonomy, human rights, democracy and the rule of law;*

The solution to this antithetical announcement goes as easy as it comes, at least in words, in the operative part, Article 1 (1) of the Convention:

> *The provisions of this Convention aim to ensure that activities within the lifecycle of artificial intelligence systems are fully consistent with human rights, democracy and the rule of law.*

Since 2024, the international community could be satisfied that there is no longer legal vacuum on AI.

An obvious merit of the Convention for the researcher is that it offers a definition of artificial intelligence systems; this would spare the author and the readers the effort of consulting a literature abundant in personal perspectives and often biased and inconsistent. Article 2:

DOI: 10.47556/B.OUTLOOK2025.23.20

> *For the purposes of this Convention, "artificial intelligence system" means 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.*

Very importantly, the Convention implies the responsibility of both public authorities or private actors acting on their behalf (Article 3a).

As we may understand, private actors that are not acting on behalf of the public authorities are not concerned by the Convention. Or, how many companies that produce AI systems, spread or sell them, are doing it on behalf of the public authorities? The drafters of the Convention were very cautious about intruding on the interests of the industry. The Convention does not apply to "activities within the lifecycle of artificial intelligence systems related to the protection of its national security interests", nor "to research and development activities regarding artificial intelligence systems not yet made available for use" (Article 3.2 and 3.3.).

With these areas and actors taken out of the purview of the Convention, what is left for regulation? The Convention specifies two general obligations: protection of human rights (Article 4) and maintaining the integrity of democratic processes and respect for the rule of law (Article 5). Yet, these obligations are already part of a solid corpus of international law, and they have an erga omnes vocation. The Convention does not bring anything new, but just states the obvious. The notion of artificial intelligence systems can be replaced with blockchain applications, drones, satellite systems, or cloud seeding devices.

The principles stipulated by the Convention do not come with anything that would deal with issues that we have identified as fundamental in defining ethics: good and bad, right or wrong, etc., pertaining to AI forms of manifestation. Those principles (human dignity and individual autonomy, transparency and oversight, accountability and responsibility, equality and non-discrimination) have to be respected "in relation to activities within the lifecycle of artificial intelligence system". As some of these principles are already embedded in language of hard international law, the Convention, as praiseworthy as it may be, does not bring new lights and does not address the main existential and epistemological worries, or about the human condition in general. The risk management measures provided by the Convention (Article 16) are abstract and vague, and circumscribed by words of caution: "graduated", "differentiated", "where appropriate".

The only reference to ethics appears only once in the Convention: this recommends public discussion and multi-stakeholder consultation on implications of the artificial intelligence systems, including the ethical ones (Article 19).

We should therefore look for more content, in other attempts to deal with ethics.

DOI: 10.47556/B.OUTLOOK2025.23.20

The EU Artificial Intelligence Act

Another "first" is the European Union's Artificial Intelligence Act, also known as the "EU AI Act", the first comprehensive horizontal legal framework for the regulation of AI systems across the EU (EU, 2024).

Regulation (EU) 2024/1689 is radically different from the Council of Europe Convention on Artificial Intelligence as it deals with the entire complexity of artificial intelligence by laying down a uniform legal framework. In particular it relates to the development, the placing on the market, the putting into service and the use of artificial intelligence systems in the Union, in accordance with Union values, to promote the uptake of human centric and trustworthy artificial intelligence while ensuring a high level of protection of health, safety, fundamental rights as enshrined in the Charter of Fundamental Rights of the European Union, including democracy, the rule of law and environmental protection.

The key goals include: a) harmonisation of rules (by establishing a uniform legal framework for the development, marketing, and use of AI systems across the EU and preventing fragmentation due to divergent national regulations); b) promotion of trustworthy AI (aspiring to the uptake of human-centric and trustworthy AI technologies that align with EU values); c) protection against potential harmful effects of AI systems, for users and society at large.

The EU AI Act develops a risk-based approach, by introducing a risk classification for AI systems, categorising them based on their potential impact and imposing specific obligations accordingly. Notably, the EU AI Act offers the same two-sentence definition of an AI system as the Council of Europe, merged into a single sentence (Article 3(1)):

> *"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".*

Unlike the Council of Europe Convention, the EU AI Act concerns and defines all actors involved in the production and marketing of AI systems and defines them in legal terms.

The EU AI Act does not explicitly use the terms "ethics" or "ethical" in its text, but it incorporates ethical considerations throughout its framework. The Act emphasises principles that align with ethical standards, such as human rights, transparency, and accountability. The EU AI Act promotes a human-centric approach to artificial intelligence, aspiring to ensure that AI systems respect fundamental rights and European values. This approach is rooted in the belief that human dignity and autonomy must be central to AI development and deployment.

DOI: 10.47556/B.OUTLOOK2025.23.20

This risk-based approach ensures that ethical considerations are integrated into the design and implementation of AI technologies.

The EU Act categorises AI systems into different risk levels, unacceptable, high-risk, and low-risk, each with corresponding regulatory requirements. Unacceptable AI practices, such as those that manipulate human behaviour or violate privacy rights, are outright banned. High-risk systems must comply with stringent standards that include transparency, accountability, and human oversight. This risk-based approach ensures that ethical considerations are integrated into the design and implementation of AI technologies.

A significant focus of the Act is placed on transparency. It mandates that users be informed when they are interacting with an AI system rather than a human. High-risk systems must also explain their decision-making processes, allowing users to understand how outcomes are derived. This requirement aims to foster trust in AI technologies by ensuring accountability.

Even if the Act itself does not make direct reference to "ethics", it is closely tied to the broader context of ethical guidelines established by the EU, known as the EU Ethics Guidelines for Trustworthy AI (EP, 2019). These guidelines advocate for a human-centric approach to AI that is lawful, ethical, and robust, ensuring adherence to fundamental rights and values. The key ethical requirements outlined in these guidelines include human agency, oversight, robustness, safety, privacy, data governance, transparency, and fairness.

The guidelines are addressed to all AI stakeholders designing, developing, deploying, implementing, using or being affected by AI in the EU. This includes companies, researchers, public services, government agencies, institutions, civil society organisations, individuals, workers and consumers.

DOI: 10.47556/B.OUTLOOK2025.23.20

Human agency and oversight – the first key requirement for achieving trustworthy AI

- developers and users should make sure that an AI system does not hamper EU fundamental rights; a fundamental rights impact assessment should be undertaken prior to its development. Mechanisms should be put in place afterwards to allow for external feedback on any potential infringement of fundamental rights.
- human agency should be ensured, i.e., users should be able to understand and interact with AI systems to a satisfactory degree. The right of end users not to be subject to a decision based solely on automated processing (when this produces a legal effect on users or significantly affects them) should be enforced in the EU.
- a machine cannot be in full control. Therefore, there should always be human oversight. Humans should always have the possibility ultimately to over-ride a decision made by a system. When designing an AI product or service, AI developers should consider the type of technical measures that should be implemented to ensure human oversight. For instance, they should provide a stop button or a procedure to abort an operation to ensure human control.

The EU AI Act explicitly enumerates several unethical practices that are deemed to pose an "unacceptable risk" and are therefore prohibited. These practices include:

- ***Manipulative techniques:*** Using AI-based systems that employ manipulative, deceptive, or subliminal techniques to influence individuals to make decisions they would not have made otherwise, particularly if this could cause significant harm to them or others.
- ***Exploitation of vulnerabilities:*** Exploiting the vulnerabilities of individuals based on their age, disability, or socio-economic status to influence their behaviour in a harmful manner.
- ***Biometric data misuse:*** Utilising biometric data to categorise individuals based on sensitive attributes such as race, political opinions, religious beliefs, sexual orientation, or other personal characteristics.
- ***Facial recognition practices:*** Creating or expanding facial recognition databases through untargeted scraping of images from the Internet or closed-circuit television footage.

DOI: 10.47556/B.OUTLOOK2025.23.20

CONCLUSIONS: CLEANING UP OUR OWN COURTYARD?

Like other visions inspired by technologies, AI crossed the border between science-fiction and reality. It did it in very turbulent way. At the level of the consumer, generative artificial intelligence invaded our computers, whether invited or not. The AI assistants intrude into many daily routines, pressing for attention. We are already victims of large-scale marketing campaigns, whose aim is to fuel curiosity and gradually to create dependence and in time monetising services we did not really ask for. At the level of the society, the fascination for AI increases, whether fuelled by fear and uncertainty or by the promise of a better and more comfortable life. The social body will assimilate AI with all its virtues and risks, and it will succumb to its charm. Human institutions will be witness to processes that will overpass their capacity to adapt and react. People will have to live with their consequences if they remain passive and disregard the trespassing of moral and ethical codes.

> ***If cheating, plagiarising, and creating false academic reputations start in education, there will be no hope of stopping some AI evils by law***

Against this background the attempts to harness the power of AI and prevent wrongdoing are welcome but never sufficient. As it was the case in general with ICTs, no one could stop the negative phenomena that proliferate in cyber space, despite good intention of governments and international organisations, their declarations, and plans of action.

However, it is important that we are aware of the problems, of the risks, of the uncertainties, and we claim the absolute necessity of ethical norms throughout the entire life cycle of AI. After all, no technology could proliferate without the massive presence of users. Raising awareness and educating billions of future users of AI systems is indeed a way to avoid or mitigate excesses and abuses. To what extent this would translate into a transfer of power from the industry to the people remains to be seen. The codification of international law, as modest and slow as it could be, is a positive and necessary step ahead.

In the meantime, while expecting international co-operation to solve the ethical deficit in using AI systems, the academic world could start cleaning its own institutions. Statistics everywhere already show that about 50% of students cheat on admissions, exams, essay writing by using AI sources, without verifiable references, or by signing shamelessly AI generated texts. Even worse, the imposture takes over aspirants to university positions and credits. I have seen texts clearly produced by AI assistants simply copied and pasted on the Internet, with a name and a photo included. If cheating, plagiarising, and creating false academic reputations start in education, there will be no hope of stopping some AI evils by law. Whatever we now call knowledge would lose its meaning.

DOI: 10.47556/B.OUTLOOK2025.23.20

REFERENCES

Council of Europe (CoE) (2024): *Framework Convention on Artificial Intelligence and Human Rights, Democracy and the Rule of Law*. Council of Europe Treaty Series – No. 225, 5.IX.2024. Available at: https://www.coe.int/en/web/artificial-intelligence/the-framework-convention-on-artificial-intelligence

Encyclopaedia Britannica (1974): Fifteenth edition, Micropaedia, 4:578:3a. [Online] Available at: https://www.britannica.com/topic/ethics-philosophy

Erwin, D. (n.d.): *Ethical Theories. Definitions & Examples*. Study.com. [Online[Available at: https://study.com/academy/lesson/ethical-theories-overview-examples.html

European Parliament (EP) (2019): *EU guidelines on ethics in artificial intelligence: Context and implementation*, European Parliamentary Research Service, PE 640.163. Available at: https://www.europarl.europa.eu/RegData/etudes/BRIE/2019/640163/EPRS_BRI(2019)640163_EN.pdf

European Union (EU) (2024): Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down *harmonised rules on artificial intelligence*, entered into force on 1 August 2024. Available at: *http://data.europa.eu/eli/reg/2024/1689/oj*

Santa Clara University (n.d.): *Ethics in Life and Business*. [Online] Available at: https://www.scu.edu/mobi/resources--tools/blog-posts/ethics-in-life-and-business/ethics-in-life-and-business.html

Spinoza, B. (1677): *Ethics*, Part IV.

UNESCO (2021): *Recommendation on the Ethics of Artificial Intelligence*. UNESCO, code SHS/BIOI/PI/2021/1. Available at: https://www.unesco.org/en/articles/recommendation-ethics-artificial-intelligence 43pp.

UNESCO (2023a): Ethical Impact Assessment: A Tool of the Recommendation on the Ethics of Artificial Intelligence. UNESCO. Available at: https://www.unesco.org/ethics-ai/en/eia.

UNESCO (2023b): Foundation models such as ChatGPT through the prism of the UNESCO Recommendation on the Ethics of Artificial Intelligence. UNESCO. [Online] Available at: https://unesdoc.unesco.org/ark:/48223/pf0000385629 doc. SHS/2023/PI/H/12.

United Nations (UN) (2003): *Declaration of Principles, Building the Information Society: a global challenge in the new Millennium*, World Summit on the Information Society, Document WSIS-03/Geneva/Doc./4-E. Available at: https://digitallibrary.un.org/record/533621?ln=en&v=pdf

United Nations (UN) (2024a): *Strengthening International Cooperation for Combating Certain Crimes Committed by Means of Information and Communications Technology Systems and for the Sharing of Evidence in Electronic Form of Serious Crimes*. UN Office on Drugs and Crime. Resolution 79/243 of the General Assembly of the United Nations, September 2024. Available at: https://www.unodc.org/unodc/cybercrime/convention/home.html

DOI: 10.47556/B.OUTLOOK2025.23.20

United Nations (UN) (2024b): *Governing AI for Humanity*. UN AI Advisory Body, Final Report, September 2024. Available at: https://www.un.org/sites/un2.un.org/files/governing_ai_for_humanity_final_report_en.pdf

United Nations (UN) (2024c): *Seizing the opportunities of safe, secure and trustworthy artificial intelligence system for sustainable development*. General Assembly, adopted on 21st March 2024. Available at: https://docs.un.org/en/A/res/78/265

United Nations System (UNS) (2022): *Principles for the Ethical Use of Artificial Intelligence in the United Nations System*. CEB, Chief Executives Board, High-Level Committee on Programmes (HLCP), Inter-Agency Working Group on Artificial Intelligence. Available at: https://unsceb.org/sites/default/files/2022-09/Principles%20for%20the%20Ethical%20Use%20of%20AI%20in%20the%20UN%20System_1.pdf

University of Wisconsin-Madison (n.d.): *Ethics in a nutshell*. University of Wisconsin-Madison, Centre for Journalism Ethics. [Online] Available at: https://ethics.sjmc.wisc.edu/resources/ethics-in-a-nutshell/

Velazquez, M., Andre, C., Shank, T. and Meyer, M.J. (2010): What is Ethics? Santa Clara Markkula Centre for applied ethics. [Online] Available at: https://www.scu.edu/ethics/ethics-resources/ethical-decision-making/what-is-ethics/

World Summit on the Information Society (WSIS) (2003): *Building confidence and security in the use of ICTs*. WSIS Plan of Action, Action Line C5.

BIOGRAPHY

Dr Petru Dumitriu is a career diplomat, specialised in United Nations (UN) matters. He earned his doctorate with a thesis on United Nations reform. He was deputy permanent representative of Romania to the UN in New York and Geneva, and Ambassador and Permanent Observer of the Council of Europe to the United Nations–Geneva. Between 2016-2020 he worked with the Joint Inspection Unit of the United Nations system as Inspector, elected by the General Assembly. He authored, among others, the JIU reports on *Knowledge Management in the United Nations system, The United Nations – private sector partnership arrangements in the context of the 2030 Agenda, Strengthening policy research uptake, and Policies and platforms in support of learning. Towards more coherence, coordination and convergence.* He received the Knowledge Management Award in 2017 and the Sustainable Development Award in 2019 for his reports. He now works as a senior fellow and lecturer at DiploFoundation, Switzerland/Malta.

 DOI: 10.47556/B.OUTLOOK2025.23.20