**UNESCO Recommendation on the Ethics of Artificial Intelligence**

Irena Barkane[[1]](#footnote-1)

# Introduction and short overview of the instrument

UNESCO adopted the Recommendation on the Ethics of Artificial Intelligence (the Recommendation) – the first global normative instrument on the ethics of artificial intelligence (AI) in November 2021.[[2]](#footnote-2) The Recommendation aims to provide a basis to make AI systems work for the good of humanity, individuals, societies and the environment and ecosystems, and to prevent harm. It provides a framework to ensure that developments align with the promotion and protection of human rights and human dignity, environmental sustainability, fairness, inclusion and gender equality. The Recommendation focuses not only on the articulation of values and principles, but also on their practical realization, via concrete policy recommendations.

Four core **values** are central to the Recommendation which lay the foundations: 1) respect, protect and promote human rights and fundamental freedoms and human dignity; 2) environment and ecosystem flourishing; 3) ensuring diversity and inclusiveness; 4) living in peaceful, just and interconnected societies.

Further, ten core **principles** lay out a human rights based approach to the ethics of AI: 1) proportionality and do not harm; 2) fairness and non-discrimination; 3) safety and security; 4) sustainability; 5) right to privacy, and data protection; 6) human oversight and determination; 7) responsibility and accountability; 8) transparency and explainability; 9) awareness and literacy; 10) multi-stakeholder and adaptive governance and collaboration.

According to the responsibility and accountability principles AI actors and Member States should respect, protect and promote human rights and fundamental freedoms, and also promote the protection of the environment and ecosystems, assuming their respective ethical and legal responsibility, in accordance with national and international law, in particular Member States’ human rights obligations, and ethical guidance throughout the life cycle of AI systems. They also provide that appropriate oversight, impact assessment, audit and due diligence mechanisms should be developed to ensure accountability for AI systems and their impact throughout their life cycle. Both technical and institutional designs should ensure auditability and traceability of AI systems in particular to address any conflicts with human rights norms and standards and threats to environmental and ecosystem well-being.

The Recommendation sets out eleven key **areas for policy action** which allow policymakers to translate the core values and principles into action: 1) ethical impact assessment; 2) ethical governance and stewardship; 3) data policy; 4) development and international cooperation; 5) environment and ecosystem; 6) gender; 7) culture; 8) education and research; 9) communication and information; 10) economy and labour; and 11) health and social well-being.

The recommendation provides that Member States and private sector companies should develop due diligence and oversight mechanisms to identify, prevent, mitigate and account for how they address the impact of AI systems on the respect for human rights, rule of law and inclusive societies. Member States, private sector companies and civil society should investigate the sociological and psychological effects of AI-based recommendations on humans in their decision-making autonomy. AI systems identified as potential risks to human rights should be broadly tested by AI actors, including in real-world conditions if needed, as part of the ethical impact assessment, before releasing them in the market.

Member States and business enterprises should implement appropriate measures to monitor all phases of an AI system life cycle, including the functioning of algorithms used for decision-making, the data, as well as AI actors involved in the process, especially in public services and where direct end-user interaction is needed, as part of ethical impact assessment. Member States’ human rights law obligations should form part of the ethical aspects of AI system assessments.

The Recommendation calls for developing, reviewing and adapting regulatory frameworks to achieve accountability and responsibility for the content and outcomes of AI systems at the different phases of their life cycle. It also calls on Member States to introduce liability frameworks or clarify the interpretation of existing frameworks to ensure the attribution of accountability for the outcomes and the functioning of AI systems. It is stated that ultimate responsibility and accountability must always lie with natural or legal persons and AI systems should not be given legal personality themselves.

Unlike other international instruments, the Recommendation includes monitoring and evaluation chapters and means for implementation in the form of a Readiness Assessment and the Ethical Impact Assessment.[[3]](#footnote-3)

# Human Rights addressed

The Recommendation is grounded on the respect, protection and promotion of human rights and underlines the obligatory character of human rights law. The Recommendation follows the international human rights framework and does not provide new interpretations thereof, including when it concerns possible adverse human rights impacts of AI technologies. Obligations under international law are referenced throughout the text, including a general clause at the end providing that this instrument is without prejudice to international law and human rights obligations.[[4]](#footnote-4)

In addition, the Recommendation explicitly mentions the United Nations Guiding Principles on Business and Human Rights. This is done in the context of asking Member States to put in place effective measures and to ensure that other stakeholders, including private sector companies, adhere to them by, among other actions, encouraging them to develop human rights, rule of law, democracy, and ethical impact assessment and due diligence tools, in line with these Guiding Principles.

The first value set out in the Recommendation calls for respecting, protecting and promoting human rights and fundamental freedoms and human dignity. The protection of human rights and dignity is the cornerstone of the Recommendation, based on the advancement of fundamental principles such as privacy, transparency and fairness.

The first principle, proportionality and do not harm, states that in the event of possible occurrence of any harm to human beings, human rights and fundamental freedoms, communities and society at large or the environment and ecosystems, the implementation of procedures for risk assessment and the adoption of measures in order to preclude the occurrence of such harm should be ensured. The choice to use AI systems and which AI method to use should be justified in the following ways: (a) the AI method chosen should be appropriate and proportional to achieve a given legitimate aim; (b) the AI method chosen should not infringe upon the foundational values set out in the Recommendation, in particular, its use must not violate or abuse human rights; and (c) the AI method should be appropriate to the context and should be based on rigorous scientific foundations. It is provided that in scenarios where decisions are understood to have an impact that is irreversible or difficult to reverse or may involve life and death decisions, final human determination should apply. It is also explicitly stated that AI systems should not be used for social scoring or mass surveillance purposes.

Privacy, a right essential to the protection of human dignity, human autonomy and human agency, must be respected, protected and promoted throughout the life cycle of AI systems. The Recommendation calls to establish adequate data protection frameworks and governance mechanisms. Algorithmic systems require adequate privacy impact assessments, which also include societal and ethical considerations of their use and an innovative use of the privacy by design approach. AI actors need to ensure that they are accountable for the design and implementation of AI systems in such a way as to ensure that personal information is protected throughout the life cycle of the AI system.

# Challenges in the adoption and implementation

The Recommendation was adopted by 193 Member States at the 41st Session of the UNESCO General Conference in November 2021. This comprehensive instrument was two years in the making and the product of the broadest global consultation process of experts, developers, and other stakeholders from all around the world. It sets the first global normative framework while giving Member States the responsibility to apply it at their level.

The main challenge for Member States now is how to apply the provisions of the Recommendation by taking appropriate steps to give effect to the principles and norms of the Recommendation. To support effective implementation of the Recommendation, UNESCO initiated a comprehensive programme and developed two instruments: the Readiness Assessment Methodology and the Ethical Impact Assessment.

In 2022, UNESCO developed the Readiness Assessment Methodology (RAM), a diagnostic tool to support governments in ensuring AI is developed and deployed ethically.[[5]](#footnote-5) It provides an assessment of a country’s legal, social, cultural, scientific, educational, technical and infrastructural AI capacities. It also indicates whether a country’s AI systems align with the values, principles and policy areas set out in UNESCO’s Recommendation. The RAM helps governments assess how robust and agile their laws, policies and institutions are in addressing AI risks. It is a diagnostic tool and the first step for targeted capacity building to strengthen institutional and human capacities in government to deal with AI.[[6]](#footnote-6)

In 2023, UNESCO unveiled the Ethical Impact Assessment.[[7]](#footnote-7) This instrument has two goals. First, to assess whether specific algorithms are aligned with the values, principles and guidance set up by the Recommendation. And second, to ensure transparency by calling for information about AI systems and the way they were developed to be available to the public. Impact Assessment tools are gaining ground to assess the true impact of AI systems. Impact assessments are provided by the draft EU AI Act for high-risk systems, and they are proposed as part of the Council of Europe’s discussion on a Convention for AI.

# Similar initiatives elsewhere

In the last decade, many national policies, other frameworks and initiatives related to the ethics and regulation of AI technologies have been elaborated by international organizations[[8]](#footnote-8), as well as by the non-governmental organizations, private sector and professional organizations[[9]](#footnote-9), and the scientific community. Many countries have also issued their strategies on AI and are developing national AI regulatory frameworks.[[10]](#footnote-10)

The European Union(the EU)is taking a lead in developing the future AI regulatory framework. On 21 April 2021, the European Commission published a long-awaited proposal for a Regulation of the European Parliament and of the Council laying down harmonised rules on artificial intelligence, known as the Artificial Intelligence Act (the draft AI Act or AI Act)[[11]](#footnote-11) – the first legal framework on AI. The EU Council provided its updated consensus draft in December 2022. In June2023, the European Parliament approved its negotiating position on the proposed AI Act. The AI Act is expected to be adopted near the end of 2023.

The Council of Europe is also striving to develop legally binding instruments in the field of AI. In 2022, the Committee on Artificial Intelligence (CAI) was established to conduct work to elaborate convention on artificial intelligence, human rights, democracy and the rule of law. In January, 2023 the first draft of the Convention was prepared by the CAI.[[12]](#footnote-12) In July 2023, a Consolidated Working Draft of the Framework Convention on Artificial Intelligence, Human Rights, Democracy and the Rule of Law was published.[[13]](#footnote-13)

# Recommendations

The Recommendation makes a strong call to national governments to establish the necessary institutional and legal frameworks to govern AI technologies and ensure they contribute to the public good. It clearly signals the end of the “self-regulatory model” that has prevailed, prioritizing commercial and geopolitical objectives over people for too long.[[14]](#footnote-14) In accordance with the Recommendation, respect, protection and promotion of human rights should be at the centre of the upcoming European and national AI regulatory frameworks.

1. Dr.iur. Irena Barkane is a researcher and lecturer at the Institute of Legal Science, Faculty of Law, University of Latvia. Her research interests and expertise lie in the fields of artificial intelligence regulation and governance, EU law, human rights, data protection, privacy, law and technology. Email: irena.barkane@lu.lv [↑](#footnote-ref-1)
2. UNESCO. (2021). UNESCO Recommendation on the Ethics of Artificial Intelligence. https://unesdoc.unesco.org/ark:/48223/pf0000381137 [↑](#footnote-ref-2)
3. UNESCO. (2021). Key Facts UNESCO’s Recommendation on the Ethics of Artificial Intelligence. https://unesdoc.unesco.org/ark:/48223/pf0000385082.page=4 [↑](#footnote-ref-3)
4. UNESCO’s Input in reply to the OHCHR report on the Human Rights Council Resolution 47/23 entitled “New and emerging digital technologies and human rights”. https://www.ohchr.org/sites/default/files/2022-03/UNESCO.pdf [↑](#footnote-ref-4)
5. UNESCO. (2023). Readiness assessment methodology: a tool of the Recommendation on the Ethics of Artificial Intelligence.

   https://unesdoc.unesco.org/ark:/48223/pf0000385198 [↑](#footnote-ref-5)
6. UNESCO. (17 July 2023). UNESCO to support more than 50 countries in designing an Ethical AI Policy this year. https://www.unesco.org/en/articles/unesco-support-more-50-countries-designing-ethical-ai-policy-year [↑](#footnote-ref-6)
7. UNESCO. (2023). Ethical impact assessment: a tool of the Recommendation on the Ethics of Artificial Intelligence. https://unesdoc.unesco.org/ark:/48223/pf0000386276 [↑](#footnote-ref-7)
8. See, e.g., OECD. (2019). Recommendation of the Council on Artificial Intelligence. <https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0449>; Council of Europe. (2020). Recommendation CM/Rec(2020)1 of the Committee of Ministers to member States on the human rights impacts of algorithmic systems. https://rm.coe.int/09000016809e1154 [↑](#footnote-ref-8)
9. See, e.g., IEEE. (2019). Ethically Aligned Design. First Edition: A Vision for Prioritizing Human Well-Being with Autonomous and Intelligent Systems. https://standards.ieee.org/wp-content/uploads/import/documents/other/ead1e-overview.pdf [↑](#footnote-ref-9)
10. See Moltzau A. (2019). National and International AI Strategies Around the World Towards 2020. <https://medium.com/@alexmoltzau/national-and-international-ai-strategies-around-the-world-towards-2020-692b22b3c303>; Klovig S. (13 April 2023).ICO responds to UK government AI regulation plans. https://www.computerweekly.com/news/365535136/ICO-responds-to-UK-government-AI-regulation-plans [↑](#footnote-ref-10)
11. European Commission. (2021). Proposal for a regulation of the European Parliament and of the Council laying down harmonised rules on artificial intelligence (Artificial Intelligence Act) and amending certain Union legislative acts. https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex%3A52021PC0206 [↑](#footnote-ref-11)
12. https://rm.coe.int/cai-2023-01-revised-zero-draft-framework-convention-public/1680aa193f [↑](#footnote-ref-12)
13. Committee on Artificial Intelligence (CAI). (2023). Consolidated Working Draft of The Framework Convention on Artificial Intelligence, Human Rights, Democracy and The Rule of Law. https://rm.coe.int/cai-2023-18-consolidated-working-draft-framework-convention/1680abde66 [↑](#footnote-ref-13)
14. UNESCO. (2021). Key Facts UNESCO’s Recommendation on the Ethics of Artificial Intelligence. https://unesdoc.unesco.org/ark:/48223/pf0000385082.page=4 [↑](#footnote-ref-14)