

**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"**

UNESCO Recommendation on the Ethics of Artificial Intelligence

1. On 23 November 2021, UNESCO's 193 Member States adopted the first global normative instrument on the ethics of artificial intelligence,¹ which addresses the concerns outlined in Human Rights Council Resolution 47/23, specifically in the field of Artificial Intelligence (AI). UNESCO's Recommendation on the Ethics of Artificial Intelligence represents a historic and unique agreement of 193 Member States on the fundamental values, principles and policies that should govern the development of this game-changing technology. It provides concrete pathways, including innovative tools, methodologies, and initiatives to ensure maximizing the positive impact of AI, while addressing the associated risks. The Recommendation is addressed to Member States, but it provides ethical guidance to all AI actors, including the private sector.
2. In line with the Secretary-General's call noted in the Resolution 47/23, the Recommendation provides that Member States are to place human rights at the centre of regulatory frameworks and legislation on the development and use of AI. The Recommendation is firmly grounded on the respect, protection and promotion of human rights and underlines the obligatory character of human rights law. This includes ensuring that human rights are not subject to trade-offs. 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.
3. It is provided that AI actors and Member States should respect, protect and promote human rights and fundamental freedoms. They should 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, including with respect to AI actors within their effective territory and control. The ethical responsibility and liability for the decisions and actions based on an AI system should always ultimately be attributable to AI actors corresponding to their role in the life cycle of the AI system.
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 the above Guiding Principles.
5. The Recommendation provides for developing appropriate oversight, impact assessment, audit and due diligence mechanisms to ensure accountability for AI systems and their impact throughout the life cycle. Both technical and institutional designs should ensure auditability and traceability of (the working of) AI systems to address any conflicts with human rights norms and standards and threats to environmental and ecosystem well-being.
6. It further 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.

¹ <https://en.unesco.org/artificial-intelligence/ethics>

7. 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.
8. Among multiple concrete policy actions, the Recommendation proposes a new model of data governance, which ensures representativeness and quality of data, promotes a better understanding of the role of data in developing secure and equitable algorithms, and sets the rules for keeping the control over the data in the hands of users, allowing them to access and delete information as needed. It also calls on Member States to ensure that appropriate safeguards for the processing of sensitive data and effective accountability schemes are devised, and to provide redress mechanisms in the event of harm. This is based on the belief that data should be owned by individuals who should have the right to delete their data. Therefore, the consent rules should be stronger, and people need to know how their data is being used.
9. Privacy is a right. The Recommendation calls to put in place appropriate safeguards to protect this right, including addressing concerns such as surveillance. It explicitly states that AI systems should not be used for social scoring or mass surveillance purposes.
10. Another crucial policy area is on gender, with specific affirmative actions and budget mechanisms. It calls for dedicating funds to finance gender-related schemes, encourage female entrepreneurship, participation and engagement in the AI field, invest in targeted programs for girls' and women in Science, Technology, Engineering, and Mathematics (STEM) and Information, Communication & Technologies (ICT) disciplines, and to ensure that national digital policies include a gender action plan. Importantly, the provisions throughout the Recommendation are aimed at eradicating gender stereotyping and ensuring that discriminatory biases are not translated into the AI systems.
11. 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. The Recommendation calls for introduction of liability frameworks or clarifying the interpretation of existing frameworks to ensure the attribution of accountability for the outcomes and the functioning of AI systems. Ultimate responsibility and accountability must always lie with natural or legal persons and AI systems should not be given legal personality themselves. To ensure this, such regulatory frameworks should be consistent with the principle of human oversight and establish a comprehensive approach focused on AI actors and the technological processes involved across the different stages of the AI system life cycle. Whenever there is harm, full transparency of the algorithms must be ensured, and the Recommendation calls for principles such as transparency, accountability and human oversight to ensure that AI remains within the control of humans.
12. There is a strong focus on developing better regulatory frameworks, ensuring redressal mechanisms and compensation when there is harm, and ensuring institutional effectiveness. However, better institutional capacities are also key. The Recommendation calls on UNESCO to assist countries in the realization of the instrument through various means. In particular, it provides for developing specific tools. Of special relevance is the *Ethical Impact Assessment (EIA)*, which aims to help Member States and other stakeholders, primarily private sector companies, evaluate the benefits and risks of AI systems, put in place risk prevention, mitigation and monitoring measures, as well as deploy redress mechanisms for those who have been adversely affected by these new technologies. The second tool is a *Readiness assessment methodology*, which aims to assist Member States in identifying their readiness status related to the Recommendation, recognizing that Member States are at different stages of readiness with respect to implementing the Recommendation. We have constituted an expert group who will be developing and advancing these tools.
13. Further, the activities on the implementation of the Recommendation include the *Group of Friends of the Recommendation*, which consists of Member States willing to help nominate and engage experts, support knowledge creation, support elaboration and piloting of the tools indicated above, to advance analytical work and provide financial and in-kind support. Implementation also includes constituting

networks of *AI Ethics Experts Without Borders (AIEB)* to ensure the availability of dedicated interdisciplinary experts in a roster attached to UNESCO for targeted implementation of the provisions of the Recommendation in Member States, particularly in the policy field, targeting developing countries; a *Global AI ethics Network (GAIN)* mobilizing some of the world's leading universities and academic centres to support AI ethics-related capacity-building, especially the Ethical Impact Assessment and Readiness Assessment methodologies; and a *Women for Ethical AI Network (W4ethicalAI)* aimed at fostering the implementation and deployment of the gender chapter of the Recommendation by engaging leading women in industry, government, science and civil society.

14. As per UNESCO's Constitution, the Recommendation is subject to regular reporting by Member States on the steps taken to implement the instrument as well as the results achieved.
15. UNESCO and ITU co-chair the High-Level Committee on Programmes Artificial Intelligence interagency working group, composed of 38 UN entities, which brings together the ethical and technological pillars of the United Nations to provide a solid foundation for current and future system-wide efforts on artificial intelligence to ensure respect for human rights and accelerate progress on the SDGs. It focuses on policy and programmatic coherence of AI initiatives within the UN system. In this context, UNESCO is working with the United Nations Office of Information and Communication Technology (OICT) and members of the working group to develop UN system wide ethical principles on the use of Artificial Intelligence, based on UNESCO's Ethics of AI Recommendation, to ensure that ethics are embedded in all stages of the AI lifecycle and to ensure the protection, promotion and respect for human rights and fundamental freedoms.

Capacity-building on AI and human rights

16. UNESCO is working to strengthen the capacities of judicial actors at the intersection of AI and the rule of law². The training addresses the legal and human rights implications of AI and discusses the use of AI in the context of administration of justice. The training was developed based on the needs identified by a survey of judicial operators in 100 countries, and was launched in December 2021.
17. UNESCO and the United Nations Institute for Training and Research (UNITAR) also launched in September 2021 an online training course for youth on AI and Human Rights³ that explores the impact of AI on freedom of expression, right to privacy, right to equality and addresses challenges like filter bubbles, echo chambers, automated content moderation, data driven biases among other challenges.
18. UNESCO is also developing a Digital Transformation and AI Competency framework for Civil Servants to identify the skills and knowledge required by Civil Servants to foster digital governance and transformation in a manner that protects and promotes Human Rights. This project is being developed as part of the UN Broadband Commission's Working Group on AI Capacity Building⁴, co-chaired by UNESCO and Nokia.

Transparency of technology companies

19. Recent years have seen heightened public debate on how the digital information landscape often fails to respect the human rights of all in the online environment or enable direct human rights violations. It remains difficult, however, to accurately assess the human rights implications of the policies and practices of tech companies. The lack of transparency into the decisions that govern the platforms that command a large portion of time spent online globally remains a serious roadblock to ensuring human

² <https://en.unesco.org/artificial-intelligence/mooc-judges>

³ <https://en.unesco.org/news/join-unesco-and-unitars-ai-and-human-rights-course>

⁴ <https://www.broadbandcommission.org/ai-capacity-building/>

rights in the digital environment. While social media companies have increasingly started to publish transparency reports, there is little consistency in what they cover and how they are presented.

20. As part of its work to develop global standards to protect human rights in the digital sphere, UNESCO has produced a set of 26-high level Principles on enhancing the transparency of internet platform companies.⁵ Greater transparency would enhance the accountability of technology companies for their operations and strengthen the respect of human rights such as freedom of expression, privacy, access to information, gender equality, safety of journalists, and the right to education, health, and association.
21. As the policy brief that sets out the transparency principles notes, “transparency is closely linked to companies’ legal and ethical obligations to respect human rights, for example as set out in the UN Guiding Principles on Business and Human Rights (‘Ruggie Principles’), and the UN Global Compact. In principle, internet companies as private entities have at least the same social obligation as other sectors which impact upon the public.”
22. Developed through a series of consultations, the transparency principles span across issues related to content and process, due diligence and redress, empowerment, commercial dimensions, personal data gathering and use, and data access. They will serve as a resource for companies, policymakers, and regulators. While all 26 principles are relevant in considering the application of the UN Guiding Principles to technology companies, the following are directly related to questions of human rights due diligence and risk assessments:

(10). Companies should be transparent as to whether they have processes to enable people to raise concerns about content, including that which appears to violate human rights or advocates incitement to violence, hostility or discrimination, as well as inaccurate content; and they should be transparent about implementation of such processes in terms of numbers and types of complaints and actions taken;

(11). Companies should be transparent about whether they conduct risk assessments for their operations, such as in contexts of upcoming elections or in countries in conflict, highlighting any serious potential threats to freedom of expression, privacy and other human rights, as well as their proposals for mitigating those threats;

(12). Companies should disclose if they have risk assessments of any algorithms whose application can have the potential to discriminate against people unfairly, and if there are any proposed mitigation measures;

(13). Companies should publish guidelines for how they will develop ethical AI processes which make consequential decisions that can impact on human rights.

23. Related to these principles, the Windhoek+30 Declaration—adopted at World Press Freedom Day 2021 and whose principles were endorsed by UNESCO’s 193 Member States at the biannual General Conference identifies greater transparency of technology companies as the first of three pillars needed to ensure information as a public good (along with supporting free, independent and pluralistic media and strengthening media and information literacy)⁶. It calls on companies to ensure transparency in relation to their human and automated systems, provide robust notice and appeals opportunities to users, conduct transparent human rights impact assessments and engage in partnerships to support information as a public good.

Open access to scientific information

24. UNESCO recognizes the relevance of human rights-based approaches in advocating for access to scientific information and research. UNESCO has been involved in several capacity building initiatives

⁵ Puddephatt, A. Letting the Sun Shine In: Transparency and Accountability in the Digital Age. 2021. Paris: UNESCO. Issue brief in the World Trends in Freedom of Expression and Media Development series. <https://unesdoc.unesco.org/ark:/48223/pf0000377231>. The principles are available in the six UN languages and Portuguese

⁶ <https://unesdoc.unesco.org/ark:/48223/pf0000378158>

in support of inclusive knowledge societies, leveraging new technological innovations and supporting principles of ‘openness’, ‘inclusiveness’, ‘access and sharing of contents, technologies and processes’ that generate information and knowledge. One such example is UNESCO’s work with the Government of Angola to formulate a framework to monitor and evaluate progress and establish benchmarks for Open Access to information. A repository for right-based Open Access data to fuel AI has also been created.

25. In support of Article 27 of the United Nations Declaration on Human Rights, the 2021 UNESCO Recommendation on Open Science sets international standards for open science practices in view of making the process and the outputs of science more inclusive, equitable, transparent and accessible with benefits for all. It identifies core principles and values for open science practices based on the rights-based, ethical, epistemological, economic, legal, political, social, multi-stakeholder and technological implications of opening science to society and broadening the principles of openness to the whole cycle of scientific research. According to the Recommendation, access to scientific information, including scientific publications, data, open software and source code, should be as open as possible with necessary restrictions justifiable on the basis of the protection of human rights, national security, confidentiality, the right to privacy and respect for human subjects of study, legal process and public order, the protection of intellectual property rights, personal information, sacred and secret indigenous knowledge, and rare, threatened or endangered species. The Recommendation further recognizes the importance of the existing international legal frameworks, in particular on intellectual property rights including the rights of scientists to their scientific productions. It also recognizes the rights of knowledge holders to receive a fair and equitable share of benefits that may arise from the utilization of their knowledge.

Open data for AI

26. Advances in digital technology and observation instruments in scientific fields have led to exponential growth of data. In advocating for open data, UNESCO is preparing a set of guidelines to advance this goal and address trust issues in curating openness of data for SDGs, using AI. The aim of the guidelines is to support Member States’ ongoing efforts to open high-quality data for SDGs and to guide the openness of data for AI technologies. The guidelines outline concrete steps that Member States can take in opening up their data, drafting an open data policy, gathering and collecting high quality data, developing open data capacities and making the data AI ready.
-