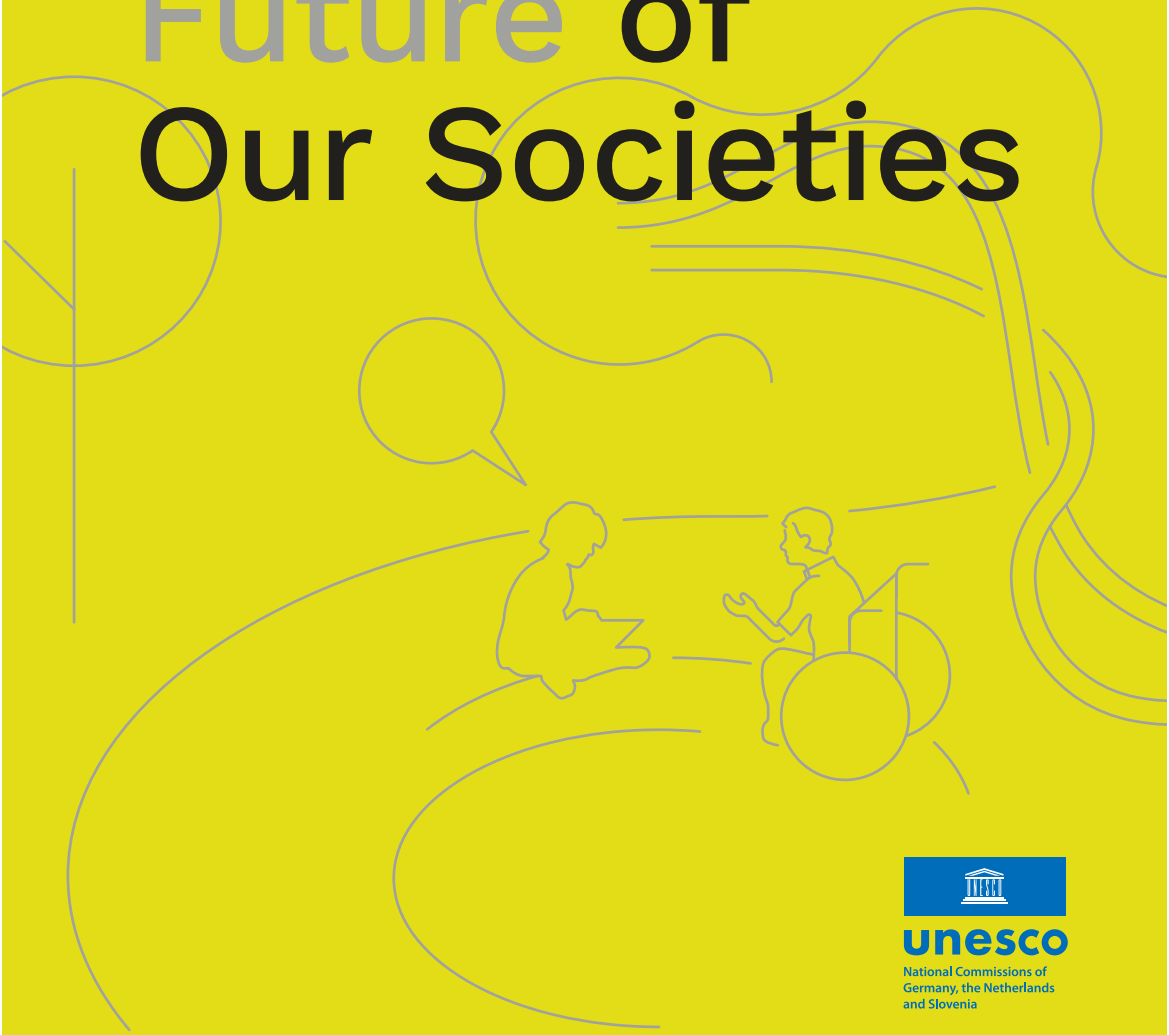
The UNESCO Recommendation on

The Ethics of AI: Shaping the Future of Our Societies



unesco
National Commissions of Germany, the Netherlands and Slovenia

The UNESCO Recommendation on The Ethics of AI: Shaping the Future of Our Societies

The UNESCO Recommendation in a nutshell

The UNESCO Recommendation on the Ethics of Artificial Intelligence (AI) is an internationally agreed legal text to ensure that any organisation, company or individual person who develops or implements AI does so in an ethical way and in line with human rights;

Ethics in AI means making sure that our interaction with AI systems is not harmful but contributes to peace, human dignity, sustainability, and security;

This will only be realised if children and adults all over the world are given the opportunity to achieve an appropriate level of AI literacy;

The Recommendation therefore acts as a guideline to all governments for creating laws and strategies on AI;

UNESCO regularly issues Recommendations. They are normative instruments of which each word is agreed upon by all 193 UNESCO Member States;

Once a Recommendation is adopted, Member States need to report back to the global community on what they have done to make sure that the agreed policy is implemented in their country.

Why do we need a global agreement on human-centric AI?

→ Artificial Intelligence has become part of our daily lives. It has enormous potential for social good, and can contribute to achieving the Sustainable Development Goals. For example, AI can be employed to predict crop yield, thereby helping to ensure food security. But the use of AI also carries risks. In several instances, AI systems have raised ethical concerns – because they may contribute to widening gender gaps, impact on people's privacy, or facilitate manipulation of media content. Since all countries are facing these issues and, at the same time, almost all AI systems are deployed on a global scale, we need a cross-border perspective on the ethical development and use of AI. In November 2021, the 193 UNESCO Member States adopted the first global agreement on human-centric AI, the Recommendation on the Ethics of AI.

What are the Recommendation's fundamental objectives?

→ The Recommendation is designed as a framework for AI systems to serve the good of humanity and the environment and to prevent harm. It is a globally accepted legal text that articulates values and principles, and how to put them into practice. It does so by suggesting concrete tasks for action to the UNESCO Member States in eleven policy areas, including health, environment, and gender. Through this approach, the Recommendation is aimed at protecting and promoting human rights and fundamental freedoms as well as human dignity and equality. Ethical questions regarding AI systems are relevant to all stages of an AI system's life cycle: from research, design and development to rollout and use, including maintenance, operation, monitoring and evaluation, and termination.

What sets the Recommendation apart?

→ The Recommendation was approved by consensus by all 193 UNESCO Member States. It places a special emphasis on the needs and contributions of low- and middle-income countries (LMICs) for ethical AI development and use. At the same time, the development of human-centric AI is defined as a task for society as a whole, going far beyond the mere regulation of technology. The Recommendation emphasises that respect for and promotion of human rights must be the basis of any AI regulation.

→ The Recommendation is characterised by the fact that it comprehensively addresses all areas of human life that are influenced by AI. In doing so, it focuses in particular on areas that have been neglected in discussions about AI up to now, **such as ...**

... different development levels of countries,

... the needs of the Global South,

... the interests of marginalised/ vulnerable groups,

... gender bias and discrimination,

... and sustainability.

Recommendation on the Ethics of Artificial Intelligence, adopted on 23 November 2021

What is 'AI'?

→ Artificial Intelligence systems are able to process data and information in a way that resembles or appears to imitate human thought processes, including aspects of reasoning, learning, perception, prediction, planning, or control.

They operate, in part, through what is called machine learning and machine reasoning.

What are the Recommendation's key requirements?

To ensure ethical AI development and use, the Recommendation establishes requirements in eleven policy areas. The following examples represent a selection of its provisions, demonstrating the scope of the Recommendation:

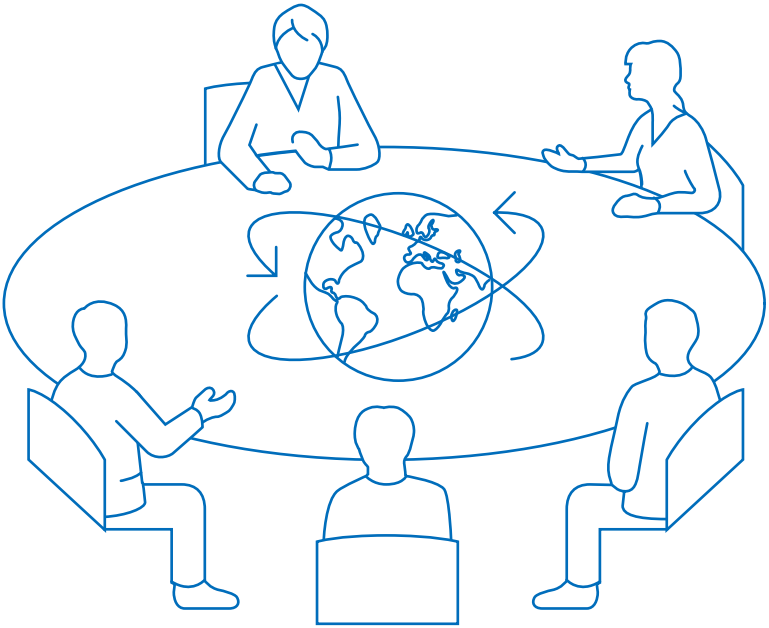
1. Laws and data governance strategies must protect the right to privacy, address concerns such as (AI-based) surveillance and ensure that individuals retain rights over their data.
2. International collaboration and cooperation on AI must be intensified to tackle urgent development challenges and to reduce the digital divide between the Global South and the Global North.
3. Ethical Impact Assessments of AI systems must be undertaken throughout their whole life cycle to prevent harm to humans and secure the promotion of human rights through AI.
4. Governance mechanisms, such as appointing AI Ethics Officers, must be in place to investigate potential harms caused through AI systems, enact enforcement and ensure redress and remedial actions.
5. Incentives must be introduced to develop AI-assisted applications for the monitoring, protection and regeneration of the environment and ecosystems.
6. AI must be developed and used in such a way that it maximises benefits for diversity and inclusiveness, including cultural diversity, safeguarding non-discrimination, promoting freedom of expression, and achieving gender equality.

7. Policies must be aimed at providing adequate AI literacy education to the public in order to empower people and to reduce digital divides.

8. AI systems must improve access to information and knowledge.

9. Governments and private sector companies must ensure a fair transition for employees whose jobs are at risk due to new AI technologies and systems.

10. Guidelines for human-robot interactions and their impact on human-human relationships must be developed, particularly bearing in mind the mental and physical health of human beings.



Calls for action in five key policy areas

The Recommendation sets out a clear ethical framework for national governments to develop concrete policies for implementation. In this translation process into national hard and soft law, both civil society and the private sector have a crucial role to play. The following paragraphs highlight five of the Recommendation's eleven policy areas of overarching societal relevance. Within these areas, concrete mandates for policies to achieve ethical development and use of AI are presented.

Health

In health care, the potential risks and opportunities of AI are more tangible than in other areas. AI already revolutionises the early detection of diseases, opens up new treatment options, and might be an opportunity to address increasing staff shortages in the (health-)care sector. However, the sector also faces important ethical issues, such as protecting personal health data and dealing with brain-computer interfaces. Administrative staff in health institutions, patients and doctors will be directly affected by these issues.

Governments and stakeholders in health care are called upon to work on the following issues:

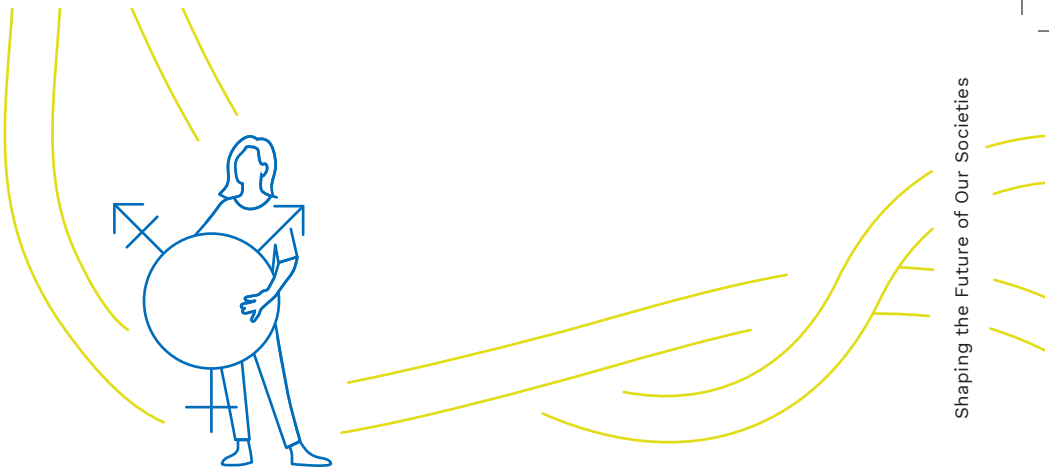
- Developing guidelines for human-robot interactions and their impact on human-human relationships, with special attention to the mental and physical health of human beings such that these human-robot interactions comply with human rights, fundamental freedoms, the promotion of diversity, and the protection of vulnerable people; especially when using AI-assisted systems for neurotechnologies and brain-computer interfaces;
- Employing AI systems to serve human health by monitoring disease outbreaks and to share data on a global scale;
- Promoting collaborative research into the effects of long-term interaction of people with AI systems, paying particular attention to the psychological and cognitive impact that these systems can have on children and young people.

Culture

AI has significant impact on how we 'produce' culture, how (much) consumers pay for cultural goods, and how artists can make a living from their work. Algorithms already influence and decide our 'consumption' of culture, even shaping, for instance, the music we listen to.

Some key responsibilities for UNESCO Member States are:

- Engaging with technology companies and other stakeholders to promote a diverse supply of, and plural access to, cultural expressions, while ensuring the visibility and discoverability of local content;
- Fostering new research at the intersection between AI and intellectual property (IP) to determine if and how AI(-assisted) works should be protected by IP rights;
- Addressing the impact of AI systems – especially natural language processing applications such as automated translation and voice assistants – on the nuances of human language and cultural expression.



Gender

The application of AI systems may introduce or reinforce bias. When organisations use translation software, it might translate gender-neutral input ('the doctor') into gendered output ('el doctor'). Similarly, when companies in a male-dominated industry recruit new candidates based on datasets with existing CVs, they will most likely reproduce a bias against female candidates. An important principle in the Recommendation is minimising and avoiding the reinforcement or perpetuation of discriminatory or biased applications and outcomes throughout the life cycle of AI systems.

Policymakers, employers in AI research and development, and other AI actors should work towards:

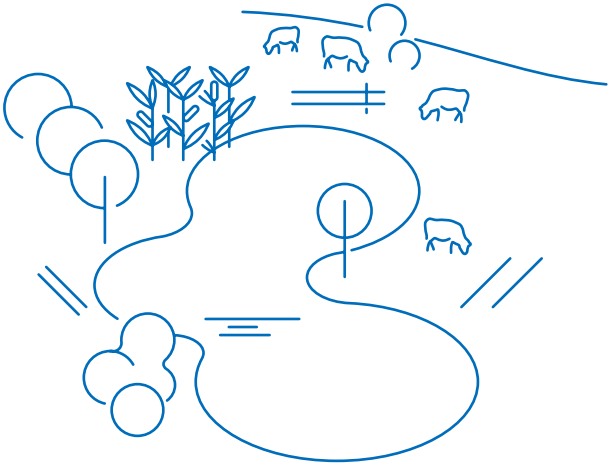
- Including a gender action plan in national digital policies;
- Increasing investments in key policy areas such as labour education to, for instance, promote women's career development in STEM and ICT disciplines;
- Ensuring that gender stereotyping and discriminatory biases are not translated into AI systems (e.g. by helping to create diverse, open and trustworthy 'gold standard datasets').

Education and Research

Schools and out-of-school learning environments are crucial for promoting awareness of the ethics of AI, and part of the effort to close the digital divide. The Recommendation presents education as an important vehicle for making sure children and adults all over the world develop AI literacy – which needs to cover both the technological and ethical dimensions of AI. At the same time, the application of e-learning in schools presents its own ethical dilemmas. For example, the use of students' personal data should comply with relevant personal data protection standards.

Member States, but also employers in education, teachers, and NGOs are urged to:

- Promote general awareness programmes about AI developments and their impact on human rights for children;
- Encourage cross-collaboration between technical skills education and AI ethics education, and promote interdisciplinary AI research by including disciplines other than science, technology, engineering and mathematics (STEM);
- Support science-policy advice to AI policy development and help to cultivate awareness of the ethics of AI.



Environment and Ecosystems

AI already contributes to protecting the environment. For example, Artificial Intelligence is used to monitor and improve biodiversity through the analysis of drone and satellite images. However, AI systems can have negative effects on the environment as well, for instance as a result of AI's significant energy demand for its computing power.

The Recommendation calls for action, among other things, on the following issues:

- Assessing the direct and indirect environmental impact throughout the AI system's life cycle and ensuring compliance with environmental regulations;
- Ensuring the development and adoption of rights-based and ethical AI-assisted applications for disaster risk resilience; the monitoring, protection and regeneration of the environment and ecosystems, and the preservation of the planet;
- Prohibiting AI systems that carry a high probability for disproportionate negative impacts on the environment.

Time to support an ethical and human-centric AI development

UNESCO's Recommendation aims to ensure that AI is developed for the benefit of societies. To make sure human-centric AI development becomes and remains the norm, input is needed from all stakeholders: not just policymakers, but also civil society organisations, private enterprises, and the general public. Despite calls for concrete action, the Recommendation underlines that a collective effort is needed. All AI actors, not just those working in health care, the culture sector, or education, should take this opportunity to co-create systems that ensure the ethical and human-centric development and employment of AI. There is work to be done!

Imprint

Published by

German Commission for UNESCO e.V.
Martin-Luther-Allee 42
53175 Bonn, Germany

Netherlands National Commission
for UNESCO
Prins Willem-Alexanderhof 20
2595 BE Den Haag, Netherlands

Slovenian National Commission for UNESCO
Kotnikova 38
1000 Ljubljana, Slovenia

Contact

German Commission for UNESCO e.V.
T +49 (0) 228 60 497-165
E zukunft@unesco.de
www.unesco.de

Netherlands National Commission for
UNESCO
T +31 (0) 70-3315484
info@unesco.nl
www.unesco.nl

Slovenian National Commission for UNESCO
T +386 (0) 1-4784799
unesco.mizs@gov.si
www.gov.si/en/state-authorities/
ministries/ministry-of-higher-education-
science-and-innovation/about-the-ministry/
urad-za-unesco/

Editors

Anne Diessner, Jeannine Hausmann
(responsible), Maximilian Müngersdorff,
Davor Orlic, Jon Verriet

Status

April 2023

ISBN

978-3-947675-35-7

Design and Layout

Panatom

supported by

Federal Foreign Office

with support of

unesco

Centre
Under the auspices
of UNESCO

IRC AI

International Research Centre
of Artificial Inteligence
under the auspices of UNESCO