



unesco

Recommendation on the Ethics of Artificial Intelligence

Maximising the benefits - while minimising the risks related to the development, adoption and deployment of Artificial Intelligence (AI) - requires an ethical framework. UNESCO, with its long track record on the ethics of science (e.g., the human genome and neuroscience) delivered the first **global instrument** on the Ethics of AI, adopted by its General Conference in November 2021.

Background

AI is already augmenting or replacing human efforts to tackle complex challenges. It generated some of the first alerts on the COVID-19 outbreak, helped accelerate the development of vaccines, and plays an increasingly important role in the recovery. However, advancements carry risks for individuals and societies. In some cases, unintended consequences of AI development and use arise, such as manipulating behaviours, greater inequalities and biases, extremism and radicalisation, or discriminating job markets. In other cases, AI is deployed with an a priori malicious intent, for example to spread disinformation or hijack core processes.

Aim and Scope

UNESCO's **Recommendation on the Ethics of Artificial Intelligence** provides a policy-friendly blueprint for the ethical governance of AI. It protects and promotes existing human rights anchored in binding international laws, and anticipates moral obligations yet to be codified.

The Recommendation covers all AI areas falling within UNESCO's mandate, and all stages of the AI life cycle: from research, design and development to deployment and use, including maintenance, operation, trade, financing, monitoring and evaluation, validation, end of use, disassembly and termination.

It takes a global approach and addresses the needs of countries at different development levels. The Recommendation's **values, principles and policy areas** are directed at promoting trustworthiness in all AI systems and throughout all AI lifecycle stages.

Universal Values

- 01 Respect, protection and promotion of human rights, fundamental freedoms and human dignity
- 02 Ensuring diversity and inclusiveness
- 03 Environment and ecosystem flourishing
- 04 Living in peaceful, just and interconnected societies

Principles

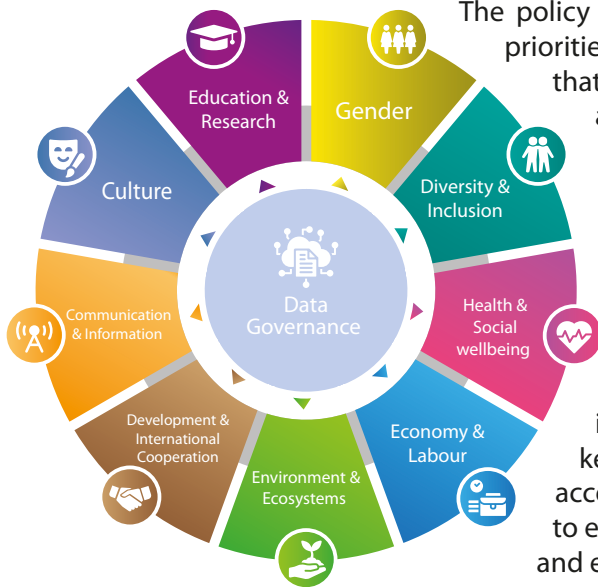
1. Proportionality and do no harm
2. Safety and security
3. Fairness and non-discrimination
4. Sustainability
5. Right to privacy and data protection
6. Human oversight and determination
7. Transparency and explainability
8. Responsibility and accountability
9. Awareness and literacy
10. Multi-stakeholder and adaptive governance and collaboration

Did you know...?

It is estimated that training one large deep-learning model produces about 300 000 kg of carbon dioxide which is equal to 125 round-trip flights between New York and Beijing. The policy area on environment calls on Member States to account for the direct and indirect environmental impact of AI systems throughout their life cycle and the data infrastructure; and to mitigate their negative impact.

Turning Values and Principles into Policy

Going beyond the “what” of values and principles, the Recommendation provides clear guidance on the “how”, to move forward and achieve impact, by proposing **policy action areas** related to:



The policy action area on **gender** puts the focus on one of UNESCO’s global priorities and on the need to increase diversity in the AI field and ensure that women and girls are represented in the development of these tools and are benefitting equally from the technology. It is designed to address a critical gap in gender and AI and to help address biases and **stereotypes in algorithms**.

Gender biases and stereotypes are also often present in AI datasets. They can help spread and fuel harmful gender stereotypes and risk further stigmatizing and marginalizing women.

The Recommendation highlights the importance of **data governance**, promotes a better understanding of the role of data in developing secure and equitable algorithms, and sets the rules for keeping control of 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.

The Recommendation underlines the need for Member States to equip workers with the **skills** needed to adjust to technological changes and thrive in the digital era, including through upskilling and reskilling. It further calls on Member States to ensure competitive markets and consumer protection, and to prevent abuse of dominant market positions and monopolies.

Tools of the Recommendation

The policy action areas in the Recommendation include **two tools**, to be developed by UNESCO, which address the challenges related to the impossibility of codifying and regulating every single aspect of AI without stifling innovation. They are motivated by the fact that, as in any innovative activity, downside risks may occur at any stage of the AI lifecycle.

Tool 1: Ethical Impact Assessment (EIA)

EIAs will have agile design and universal relevance, helping AI stakeholders assess the impact that datasets, conceptual frameworks and algorithms may have on society. This tool targets AI system developers and aims to support adjusting algorithms to address or mitigate possible negative impacts.

Tool 2: Readiness Assessment Methodology

The readiness assessment methodology will assist Member States in identifying their status related to the Recommendation. It recognizes that Member States are at different levels of preparedness with respect to developing, adopting and using AI technologies, and aims to help governments fine-tune regulatory mechanisms.

The implementation of the Recommendation

It is complemented by a number of initiatives and networks:

- ▲ **High-Level Expert Group on the Implementation of the Recommendation** to assist in developing innovative tools and methodologies;
- ▲ **Global Forum on the Ethics of AI** as a central point to debate the challenges of these technologies;
- ▲ **Global Observatory of the Ethics of AI** as a digital platform to inform about policies in this domain and advances on the ground;
- ▲ **AI Ethics Experts Without Borders** for deployment in Member States, on a needs basis, to assist in implementation, particularly in developing countries;
- ▲ **Global Network of Knowledge Centres on Ethics of AI**, mobilizing the world’s leading universities and academic centres to support capacity-building and research on ethics of AI;
- ▲ **Women for Ethical AI Network**, mobilizing prominent women champions of gender equality for the implementation of the ground-breaking gender policy area of the Recommendation.



Social and Human Sciences Sector

7, Place de Fontenoy
75352 Paris 07 SP, France

✉ g.ramos@unesco.org

🐦 @UNESCO

📷 @unesco



on.unesco.org/EthicsAI