



# Making AI Ethical by Design: The UNESCO Perspective

Gabriela Ramos, Mariagrazia Squicciarini, and Eleonora Lamm , UNESCO

*The rapid development of artificial intelligence (AI) technologies has generated ethical concerns around the globe due to the consequences of the misuse of these technologies. This article discusses the UNESCO Recommendation on the ethics of AI and outlines some key elements and their potential impacts.*

**A**rtificial intelligence (AI) is impacting all aspects of our lives, including our social, economic, political,<sup>1</sup> and cultural relationships and the way we interact with the environment.

## INTRODUCTION

Powerful new developments such as generative AI<sup>2</sup> promise to raise productivity and transform industries, for example, by automating routine tasks, speeding up drug discovery, or improving patient diagnosis.

However, these technologies may disrupt labor markets, by changing tasks or triggering job displacements. A recent study suggests that foundation AI models could impact 80% of the U.S. workforce, affecting approximately 10% of work tasks.<sup>3</sup>

These developments raise profound ethical concerns as AI systems may embed or amplify biases,<sup>4</sup> trigger redundancies through the automation of routine and nonroutine tasks,<sup>5</sup> contribute to climate degradation, and affect human rights and fundamental freedoms. These risks<sup>6</sup> have already emerged and have begun to add to existing inequalities, at times resulting in harming already marginalized groups.<sup>7</sup>

This calls for coordinated action supported by global frameworks and regulations that allow individuals,

Digital Object Identifier 10.1109/MC.2023.3325949  
Date of current version: 7 February 2024

societies, and economies alike to leverage the possibilities offered by AI while addressing its drawbacks. The United Nations Educational, Scientific and Cultural Organization (UNESCO) Recommendation on the Ethics of Artificial Intelligence (hereafter the Recommendation),<sup>8</sup> adopted in 2021 by UNESCO's (at that time; the United States rejoined UNESCO in July 2023) 193 member states, does just that. It aims to serve as a compass for the design, development, deployment, and procurement of AI systems that respect human rights and human dignity and preserve fundamental freedoms and the environment.

It is vital to ensure that the growth and development AI may bring are equitably distributed among and within nations and that they are not concentrated in the hands of a few companies or countries. To this end, governments need to take steps to ensure competitive markets and consumer protection, avoid abuse of dominant market positions and monopolies, and ensure that people are not left behind. Providing strong ethical guardrails is essential also to safeguard investment in AI and make AI technologies transparent, explainable, accountable, and fair. This will in turn help enhance trust in AI and contribute to leveling the playing field, benefiting people while allowing small and medium enterprises to participate and compete in the AI arena. Being ethical is good for business as ethical companies attract customers who value responsible and trustworthy AI solutions.<sup>9</sup>

The recent global debate demonstrates<sup>10</sup> that regulating a technology as powerful and potentially divisive as AI is no longer taboo. The question is not whether to regulate but how to best do so, to allow technology to flourish

and help solve societal problems while containing its possible adverse effects, ensuring that everybody can benefit from it.

In what follows, this article discusses how the Recommendation can help address the concerns posed by AI and outlines some of the key elements put forward and their potential impacts on different sectors. It then outlines the way UNESCO is implementing the Recommendation and the tools developed for the purpose, namely, the readiness assessment methodology (RAM) and the ethical impact assessment. Before concluding, this article discusses what regulating AI may entail in a view to maximizing societal benefits.

### THE UNESCO RECOMMENDATION ON THE ETHICS OF AI

UNESCO is a specialized agency of the United Nations having the mandate to promote peace through education, culture, and science, both natural and social and human sciences (SHSs). The work undertaken in relation to SHSs aims to promote inclusive, fair, and sustainable societies. This, among others, is pursued through standard-setting activities and capacity-building programs fostering the ethical development and deployment of science and new technologies, such as biotechnology, neurotechnology, and AI.

Following the request of its member states, in 2019, UNESCO took up the challenge to develop the first-ever global standard on the ethics of AI, in the form of a recommendation,<sup>8</sup> with the aim to translate the *what* of AI ethics into the *how* of relevant policy action. The Recommendation moves beyond high-level principles,

developing practical and implementable strategies<sup>11</sup> and translating its values and principles into eleven key areas for policy actions. These encompass data governance, environment and ecosystem protection, gender inclusion, education and research, and health and social well-being, among others.

In what follows, this article focuses on the areas that are most relevant to the computer community and to the development and deployment of AI systems that are ethical by design.

### DATA POLICY

The Recommendation emphasizes the need to ensure effective data governance,<sup>12</sup> including the organizational frameworks responsible for overseeing data management, storage, processing, analysis, utilization, sharing, and transactions carried out directly by or on behalf of governments.<sup>13</sup> Data governance approaches vary depending on laws and social structures and may be influenced by the roles and actors active in the data ecosystem.<sup>14</sup>

The accelerated pace at which the digital transformation has been unfolding recently has contributed to highlighting the role of data and their use. A case in point is the COVID-19 pandemic, during which, while rushing to implement life-saving applications, the ethical governance of data was at times overlooked. This led to abuses, a lack of transparency, and asymmetries of power,<sup>15</sup> triggering concerns regarding data privacy, data security, and data governance.<sup>16</sup>

According to a survey conducted in Australia, 58% of respondents admitted that they do not understand how companies utilize the data they gather, and 49% felt helpless with respect to

protecting their data due to a lack of knowledge or time or the inherent complexity of the processes.<sup>17</sup>

The Recommendation acknowledges these power imbalances and the lack of understanding from data constituencies, highlighting the need to balance individuals' privacy while facilitating data collection and regulating its use. It calls for high-quality, representative, and robust datasets to train, develop, and deploy AI systems.

Today, data are collected both directly and indirectly, and sensors embedded in different devices and systems enable the collection of vast amounts of data from the physical world (for example, temperature and location), thus multiplying data-rich environments. While this can enhance our ability to map and report on the world around us,<sup>18</sup> it can make individuals' acts or movements fully traceable, for good or bad purposes. Moreover, while data analytics techniques, for example, deep learning and neural networks, help uncover patterns and relationships within data, thus generating valuable knowledge in support of decision-making, decisions relying on insufficient, partial, or low-quality data compromise the accuracy and reliability of the proposed solutions.<sup>19</sup>

Additionally, ensuring the representation of different groups in relation to, for example, race, gender, age, socioeconomic, or disability status when training AI datasets is crucial to ensure the accuracy of the AI's predictions or decisions. For example, the underrepresentation of older adults in AI data and the lack of consistent methods to collect and record age-related information may lead to solutions overlooking their needs, discriminating against them or offering inadequate solutions.<sup>20</sup>

To achieve this, the Recommendation encourages the creation of "gold standard" datasets that could serve as benchmarks for quality and accuracy of other datasets. In addition, it advocates for the availability of open and trustworthy datasets and for stakeholders to contribute to the transparency, reproducibility,

## ETHICAL GOVERNANCE AND STEWARDSHIP

Effective AI governance mechanisms are paramount. They need to be inclusive, ensuring that the benefits of AI are accessible to all members of society, and transparent, allowing stakeholders to access relevant information, in turn fostering understanding

**CONVERGING TECHNOLOGIES, THAT IS, ORIGINALLY UNRELATED TECHNOLOGIES THAT BECOME INTEGRATED AS THEY DEVELOP AND ADVANCE SUCH AS AI APPLIED TO NEUROTECHNOLOGY, CALL FOR CAUTIOUS DATA GOVERNANCE.**

and accountability of AI research and development. For instance, it calls for a "data trust" as a mechanism that can facilitate data flows and the exchange of transparent, accountable, and trustworthy data.<sup>21</sup>

### Converging technologies

Converging technologies, that is, originally unrelated technologies that become integrated as they develop and advance such as AI applied to neurotechnology, call for cautious data governance. Brain data collected by tracking our brains, reactions, and so on are delicate and at risk of manipulation. That is why the Recommendation includes provisions focusing on neurotechnology (see "Neurotechnologies") and brain-computer interfaces, with the aim to safeguard human dignity and autonomy<sup>22</sup> and address the potential manipulation and exploitation of human cognitive biases.

and trust in AI. Multidisciplinary approaches are equally essential as they allow for the integration of different disciplines in the governance of AI.

AI governance should also be multilateral, to help ensure a comprehensive global perspective, accounting for the needs and contexts of different regions. To achieve this, a multistakeholder approach is necessary<sup>23</sup> ensuring that AI systems reflect a diverse range of perspectives.

Governance actions should further be embedded in institutional capacities. An ability to anticipate potential issues should be coupled with robust enforcement and redress mechanisms. The Recommendation stresses that AI governance should identify risks, impacts, and harms and that AI governance should then proceed with investigations and concrete remedial actions.

## NEUROTECHNOLOGIES

**N**eurotechnology refers to devices and procedures used to access, monitor, investigate, assess, manipulate, and emulate the structure and function of neural systems. These include 1) technical and computational tools that measure and analyze chemical and electrical signals in the nervous system and 2) technical tools that interact with the nervous system to change its activity.<sup>S1</sup>

Recently, markets have witnessed an increased availability of neurotechnology-related consumer products to be used for recreational or mental augmentation purposes. The effects of these technologies are still unclear, and even if they are regulated when related to medical, invasive procedures, this is not necessarily the case when they converge with AI and in human-computer interactions.<sup>S2</sup>

Neurotechnology stands out due to its direct interface with and influence over the brain, which carries profound implications for human existence,<sup>22</sup> including

- » mental integrity, that is, individuals' control over their mental states and brain data, ensuring that no one can read, alter, or spread such information without consent
- » autonomy and free decision-making, where the monitoring, surveillance, and manipulation of cognitive functions that neurotech facilitate may interfere with a person's capacity for self-directed actions and decision-making
- » mental privacy as neurodevices can help reveal individuals' brain activity and lead

to discriminatory practices, also related to neurological or mental health conditions.

Moreover, the neurotech world is highly concentrated at present. This triggers additional challenges related to accessing the technology and power imbalances, among others. Over 80% of high-impact neuroscience publications are found to be produced by only ten countries, while 70% of countries have contributed fewer than ten such papers over the last two decades. Similarly, six countries hold 87% of IP5 neurotech patents.<sup>22</sup> This concentration creates divides between countries that are at the forefront of neurotech and the others, possibly perpetuating or increasing disparities in health-care, research, and innovation.<sup>S3</sup>

### REFERENCES

- S1. "Report of the International Bioethics Committee of UNESCO (IBC) on the ethical issues of neurotechnology," UNESCO, Paris, France, 2021. [Online]. Available: <https://unesdoc.unesco.org/ark:/48223/pf0000378724>
- S2. UNESCO, University of Milan-Bicocca (Italy), State University of New York, and Downstate Health Sciences University, "The risks and challenges of neurotechnologies for human rights," UNESCO, Paris, France, 2023. [Online]. Available: <https://unesdoc.unesco.org/ark:/48223/pf0000384185>
- S3. A. Ponne, "Ethics of neurotechnology," UNESCO, Paris, France, n.d. [Online]. Available: <https://www.unesco.org/en/ethics-neurotech>

Moreover, as AI technologies evolve rapidly, AI governance systems must be flexible enough, also to accommodate cultural or legal differences.<sup>23</sup>

### EDUCATION AND RESEARCH

New developments such as ChatGPT and other large language models are creating high expectations for the services

they may provide. At the same time, their widespread use is laying bare the risks attached to the way these technologies are being deployed, responding

to a technological race among economic actors, both companies and countries, that seldom put the public good as the first priority.<sup>24</sup>

Preliminary assessments show that generative AI can deliver misleading, inaccurate, or false information without making this clear to the user (ChatGPT introduced a disclaimer only recently).<sup>25</sup> Authorship and intellectual property rights-related issues remain largely unaddressed as platforms do not quote their sources or are not transparent about their work.

This is why the UNESCO Recommendation asks for adequate AI literacy and education to be provided to the public, including awareness programs about data and the ethical use of AI technologies. When doing so, the participation of marginalized groups should be prioritized, to reduce digital divides and digital access inequalities. Evidence points to the need for people to have technical skills related to AI itself (that is, machine learning, Python, or neural networks), as well as socioemotional skills such as creativity, ability to work in teams, and communication for individuals to thrive in the AI era.<sup>26</sup> In addition, digital literacy emerges as a basic requirement for every citizen to be able to participate fully in the digital society.<sup>27</sup>

The Recommendation further asks member states to support research initiatives on ethical AI.

## ECONOMY AND LABOR

AI has the potential to make workers and firms more productive and efficient and to spur innovations, new products, and new services.<sup>5</sup> At the same time, AI can automate (parts of) existing jobs,<sup>28</sup> thus leading to redundancies and to exacerbating inequalities. Moreover algorithms

## SOME UNFAIR DATA

**T**wenty percent more men than women received a Facebook ad for STEM careers.<sup>S4</sup>

A 2022 study<sup>S5</sup> shows how the use of artificial intelligence (AI) by employers disadvantages women in the workplace, throughout their working lives, whereas a 2020 report finds that women make up just 26% of data and AI positions in the workforce.<sup>S6</sup> Also, women accounted for less than 19% of all AI and computer science Ph.D. graduates in North America over the past ten years.<sup>S7</sup>

Women further experience glass ceiling issues resulting in vertical gender segregation. Women in STEM fields and the digital sector are less likely to hold high-level positions. UNESCO in 2019<sup>S8</sup> found that just one in every four leadership positions in tech industries (including nontechnical positions in marketing, human resource management, and the like) was occupied by women.

### REFERENCES

- S4. A. Lambrecht and C. E. Tucker, "Algorithmic bias? An empirical study into apparent gender-based discrimination in the display of STEM career ads," *Manage. Sci.*, vol. 65, no. 7, pp. 2966–2981, 2019, doi: 10.1287/mnsc.2018.3093.
- S5. "The effects of AI on the working lives of women," UNESCO, OECD, IDB, Paris, France, 2022. [Online]. Available: <https://unesdoc.unesco.org/ark:/48223/pf0000380861>
- S6. "Global gender gap report 2020," World Economic Forum, Cologny, Switzerland, 2020. [Online]. Available: [https://www3.weforum.org/docs/WEF\\_GGGR\\_2020.pdf](https://www3.weforum.org/docs/WEF_GGGR_2020.pdf)
- S7. D. Zhang et al., "The AI index 2021 annual report," AI Index Steering Committee, Human-Centered AI Inst., Stanford Univ., Stanford, CA, USA, Mar. 2021.
- S8. "I'd blush if I could: Closing gender divides in digital skills through education," UNESCO, Paris, France, GEN/2019/EQUALS/1 REV., 2019. [Online]. Available: <https://unesdoc.unesco.org/ark:/48223/pf0000367416.page=1>.

based on nonrepresentative datasets can lead to discrimination against (some) workers.

While previous technological advances have tended to automate "routine" tasks, AI, and generative AI in

## TAKING ACTION: WOMEN FOR ETHICAL AI

**T**o fight gender stereotypes and discrimination, UNESCO launched the Women4Ethical Artificial Intelligence (AI) (W4EAI) Platform, a group of female professionals and experts promoting gender diversity at the institutional level. The platform supports establishing country-specific frameworks promoting ethical AI policies and applies a gender lens to issues such as the environment or discrimination, to ensure wide-reaching and impactful solutions. Confronting these challenges requires multistakeholder collaboration and female-led empowerment, which is the primary goal of the W4EAI Group.

The platform's work will strengthen the ability of UNESCO's member states to promote trustworthy and gender-balanced AI systems based on the Recommendation.<sup>59</sup>

### REFERENCE

- S9. "Artificial intelligence: UNESCO launches women4ethical AI expert platform to advance gender equality," UNESCO, Paris, France, Apr. 28, 2023. [Online]. Available: <https://www.unesco.org/en/articles/artificial-intelligence-unesco-launches-women4ethical-ai-expert-platform-advance-gender-equality?hub=32618>

particular, has the potential to automate "nonroutine" tasks, that is, tasks entailing creativity or requiring high levels of skills to be accomplished, thus exposing large new swaths of the workforce to potential disruption and redundancies.<sup>29</sup>

Emphasizing the importance of addressing the impact of AI systems on the labor market, the Recommendation advocates for the integration of AI-related education and training into education and training at all levels. This would help address the lack of skills needed to work and live with AI. AI-related education and training

would, in turn, help boost the labor market, spur competition, and foster consumer protection, nationally and internationally.<sup>30</sup> The Recommendation further invites member states to work with the private sector, civil society organizations, and other stakeholders, including workers and trade unions, to ensure a fair transition for at-risk employees. This includes putting in place upskilling and reskilling programs, finding effective mechanisms to retain staff during transition periods, and exploring "safety net" programs for those who cannot be retained or retrained.

## GENDER

In this context, member states felt the need to ensure that AI is gender-inclusive (see "Some Unfair Data"). UNESCO's Recommendation is the first and only instrument featuring a full policy chapter on gender and AI (see "Taking Action: Women for Ethical AI"). It calls for the collection of gender-disaggregated data (what cannot be measured cannot be fixed) and outlines concrete approaches to ensure that AI systems do not replicate gender-based disparities. It also encourages female entrepreneurship, participation, and engagement.

## ENVIRONMENT AND ECOSYSTEMS

As AI consumes a lot of energy and impacts the environment (see "Computing Consumes"), the Recommendation asks member states and businesses to assess the direct and indirect environmental impacts of AI systems and ensure that AI solutions help support the prevention, control, and mitigation of climate-related problems. These include carbon footprints, energy consumption, and raw material extraction.<sup>31</sup>

### Water consumption

A 2023 study finds that training GPT-3 in the U.S. data centers directly consumed 700,000 L of clean fresh water (equivalent to producing 370 BMW cars or 320 Tesla electric vehicles). Such water consumption would be three times as much if the training happened in Microsoft's Asian data centers.<sup>33</sup> Another 2023 study finds that ChatGPT requires 500 mL of water for every 20 to 50 questions answered, to keep temperatures in range and counter the heat triggered by the energy used in the training process.<sup>34</sup> This is concerning as freshwater scarcity is one of the most pressing challenges of humanity.



AI models should address such concerns and address their own environmental footprint,<sup>33</sup> by developing energy-efficient algorithms, optimizing computing resources, and implementing sustainable practices in AI infrastructures.<sup>35</sup> This includes designing algorithms that require less computing power or use renewable energy sources to power AI systems. These efforts can reduce the carbon footprint<sup>36</sup> of AI and contribute to climate change mitigation<sup>37</sup> and the sustainability of AI.<sup>38</sup>

## IMPLEMENTING THE RECOMMENDATION ON THE ETHICS OF AI

The current regulatory situation is one whereby on one hand, many countries, particularly in the developed world, are trying to keep pace with technological developments by drafting national or regional strategies, enacting laws, such as the AI Act in the European Union,<sup>39</sup> and building institutional frameworks to establish ethical guardrails. On the other hand, many countries are lagging behind in harnessing the power of these technologies and containing the risks associated with them.

This is a missed opportunity as AI has the potential to enable a more equitable, fair, and sustainable future.<sup>40</sup> For benefits to materialize, though, governments should promote the ethical development and use of AI and develop the relevant capacities and skills, in both the public and private sectors. They should also advance the relevant regulatory and institutional frameworks, to turn strategies into reality.

Many countries around the world are now in the process of implementing the Recommendation, with UNESCO's support, with the aim to translate it into national, institutional, and

## COMPUTING CONSUMES

- » A common AI training model can emit more than 626,000 pounds of CO<sub>2</sub> equivalent, about five times the lifetime emission of the average car, including the manufacture of the car itself.<sup>S10</sup>
- » BLOOM's (the less energy intensive of the four language models analyzed by the Stanford Institute for Human-Centred Artificial Intelligence, along with GPT-3, Gopher, and OPT) training emitted 1.4 times more carbon in one year than the average American and was the equivalent of a single passenger flying from New York to San Francisco 25 times. BLOOM's training consumed enough energy to power the average U.S. home for 41 years.<sup>32</sup>

**REFERENCE**

S10. E. Strubell, A. Ganesh, and A. McCallum, "Energy and policy considerations for deep learning in NLP," 2019. [Online]. Available: <https://arxiv.org/pdf/1906.02243.pdf>

regulatory frameworks promoting the ethical use and development of AI.

### Walking the talk

To ensure the effective implementation of the Recommendation, UNESCO developed the RAM, a practical tool assisting countries in understanding the current state and level of their preparedness to undertake the ethical and responsible implementation of AI technologies. According to Stanford University's 2023 AI Index, 37 AI-related bills were passed into law globally in 2022, with the United States alone passing nine laws.<sup>32</sup> Leveraging the RAM, countries can identify the areas where institutional and regulatory changes may be desired: without the necessary institutions, human capital, capacity building, and infrastructure,

there can be no meaningful advancement or progress.

The RAM encompasses five dimensions:

1. *Legal/regulatory*: It addresses member states' institutional and human capacity to implement the Recommendation and manage the major societal transformations that an increased adoption of AI in all sectors of the economy would trigger. This includes appropriate regulatory frameworks, as well as monitoring and evaluation mechanisms.
2. *Social/cultural*: It considers the factors relevant for the ethical development and deployment of AI, for example, inclusiveness, social and cultural

diversity, and public awareness. It also addresses attitudes to AI technologies, including public acceptance, to help reveal societal values and preferences leading societal choices.

3. *Economic*: It aims to address the size and strength of the supply side of the AI ecosystem in the country and its ability to develop AI solutions that reflect the specific needs and conditions.
4. *Scientific/educational*: It assesses the AI research and development undertaken in a country (for example, AI-related publications and patents, AI researchers and engineers involved in R&D, etc.). It also analyzes AI ethics research and the availability of educational opportunities for students and the public. It further monitors

the number of students (graduates), professionals, and public in AI, ICT, or STEM disciplines and the level of AI and e-skills in the population.

5. *Technical/infrastructural*: It assesses the level of ICT and related technical infrastructure in place and reflects the idea that without the relevant infrastructure, AI development and the implementation of AI-based solutions cannot be scaled up. UNESCO is currently engaged with more than 50 countries worldwide to deploy the RAM and will continue expanding this work, to strengthen global impact.

### ENGAGING THE PRIVATE SECTOR AND DEVELOPERS

Although the Recommendation primarily targets member states, it also

provides ethical guidance to other AI actors, including communities, institutions, and companies (see “The Business Council for Ethics of AI”). Implementing ethical frameworks leads to win-win outcomes. Society benefits from AI technology that aligns with human rights standards. Companies can differentiate themselves from competitors and attract customers valuing trustworthy AI solutions.<sup>9</sup> Evidence shows that incorporating responsible AI practices can yield competitive advantages by enhancing product quality<sup>41</sup> and can positively impact talent acquisition, retention, and engagement, particularly as employees increasingly scrutinize their employer’s ethical values. The absence of responsible AI practices within the tech sector is challenging trust and triggering branding risks.

### WHICH REGULATION?

Given the sheer impact of AI on everybody’s life, the time of light-touch self-regulation should be considered as over. Good regulation is essential to foster robust, enabling environments for responsible AI. The supposed existence of a dichotomy between regulation and innovation is flawed as effective regulatory frameworks provide certainty, and once the playing field is leveled, innovation flourishes, benefiting both users and innovators, all innovators, including small- and medium-sized firms. The question is not whether or not to regulate but how to best do so.

Many voices, both private and public, are now calling for a review of the way AI technologies are developed and deployed, signaling the need for stronger governance and oversight capacities. Since 2021, when the Recommendation on the Ethics of AI was adopted, UNESCO has been building the tools and support systems for its

## THE BUSINESS COUNCIL FOR ETHICS OF AI

**T**ogether with a number of companies such as Telefonica and Microsoft, UNESCO has cocreated a network of companies, the Business Council for Ethics of Artificial Intelligence (AI).<sup>S11</sup> The council will serve as a space to exchange experiences and good practices, strengthen technical capacities in ethics and AI, design and implement the ethical impact assessment, and ensure that AI develops in an ethical manner.

### REFERENCE

- S11. “Business council for ethics of AI,” UNESCO, Paris, France, n.d. [Online]. Available: <https://www.unesco.org/en/artificial-intelligence/ibero-america-network1>



implementation. When lacking comprehensive regulatory frameworks, countries risk being unable to assess and address responsibilities and liabilities: clear and accountable AI frameworks are paramount to secure the rule of law in the digital world.

Several countries are working on the development of AI policy and legal frameworks, including regulations. In 2022 alone, laws mentioning the term “AI” were debated in the legislative bodies of 127 countries, and 37 of them were passed (for a total of 123 laws), according to Stanford University’s AI Index Report 2023.<sup>32</sup>

The goal should not be AI regulation per se but rather building an AI-enabling environment that contributes to the public good. To this end, investment in governments’ capacities to deal with these technologies is also needed. Getting the regulations right is vital, but it is the capacities of governments and government officials, and the right institutional design, that will determine their success.

We need to avoid the abuse and misuse of AI technologies, for example, to spread hate speech and misinformation and interfere with democratic processes. We need to promote diversity and inclusion across the entire AI system life cycle, particularly by supporting the participation of women and developing countries. We need to fight against the monopolization of AI technologies to ensure equitable access and shared benefits. We need to ensure transparency when decisions are made with AI to increase public scrutiny and foster trust in AI technology.

At first glance, this can sound like an insurmountable task, but history is replete with examples of humanity coming together to address critical issues.

## ABOUT THE AUTHORS


**GABRIELA RAMOS** is the assistant director-general of the Social and Human Sciences Sector of UNESCO, where she oversees the institution’s contributions to building inclusive societies. Her mandate includes tackling economic inequalities of income and opportunity and promoting social inclusion and gender equality. Ramos received a master’s in public policy from Harvard University. She is a member of the G7 Gender Equality Advisory Council for 2022, an independent advisory body that develops recommendations on gender equality issues across the entire G7 agenda. She is a board member of the Paris Peace Forum and the UNICEF Advisory Board, as well as the Lancet Commissions on COVID and on Men’s Violence against Women, and is a member of the Danone “Entreprise à Mission” Committee. Contact her at [g.ramos@unesco.org](mailto:g.ramos@unesco.org).

**MARIAGRAZIA SQUICCIARINI** is the chief of executive office of the Social and Human Sciences Sector of UNESCO. She leads and supervises policy-relevant work related to, among others, the transformation of societies, inequalities, and inclusion in the digital era; AI and neurotechnology; and the digital gender divide. Squicciarini received a Ph.D. in economics from the University of Essex. Contact her at [mg.squicciarini@unesco.org](mailto:mg.squicciarini@unesco.org).

**ELEONORA LAMM** is the head of the Social and Human Sciences Sector for Latin America and the Caribbean (AI) at UNESCO. She leads the strategy of the implementation of UNESCO Recommendation on the Ethics of Artificial Intelligence, doing research and providing policy advice to countries. Lamm received a Ph.D. in bioethics and law from the University of Barcelona. Contact her at [e.lamm@unesco.org](mailto:e.lamm@unesco.org).

Getting AI right will be another example of what can be accomplished through a shared vision and coordinated actions. The UNESCO Recommendation in this case represents a blueprint for global consensus on the “what,” as well as the “how” of ethical AI.

In this regard, liability is important. Compensation if there is damage needs to be provided for, as well as human determination, accountability, and the nongranteeing of legal personality to AI developments.

These goals are ambitious, and to achieve them, we need the entire computer community to get involved. Computer science and social sciences cannot continue to develop in parallel, and ethics needs to accompany any scientific development. 

## REFERENCES

1. L. Bosser, M. Cantero, I. Galariotis, and D. Innerarity, “Neither fish nor fowl: The challenges of a democratic AI,” European Univ. Inst., Fiesole,

- Italy, 2022. [Online]. Available: <http://hdl.handle.net/1814/75047>
2. G. Cooper, "Examining science education in ChatGPT: An exploratory study of generative artificial intelligence," *J. Sci. Educ. Technol.*, vol. 32, no. 3, pp. 444–452, 2023, doi: 10.1007/s10956-023-10039-y.
3. T. Eloundou, S. Manning, P. Mishkin, and D. Rock, "GPTs are GPTs: An early look at the labor market impact potential of large language models." OpenAI. Accessed: Nov. 20, 2023. [Online]. Available: <https://openai.com/research/gpts-are-gpts>
4. K. Wach et al., "The dark side of generative artificial intelligence: A critical analysis of controversies and risks of ChatGPT," *Entrepreneurial Bus. Econ. Rev.*, vol. 11, no. 2, pp. 7–24, 2023, doi: 10.15678/EBER.2023.110201.
5. D. Acemoglu, D. Autor, J. Hazell, and P. Restrepo, "Artificial intelligence and jobs: Evidence from online vacancies," *J. Labor Econ.*, vol. 40, no. S1, pp. S293–S340, 2022, doi: 10.1086/718327.
6. E. Enoch, "Will ChatGPT replace your job as a programmer?" Medium. Accessed: Nov. 20, 2023. [Online]. Available: <https://medium.com/codex/will-chat-gpt-replace-your-job-as-a-programmer-3492ad2cf449>
7. J. Whittlestone and S. Clarke, "AI challenges for society and ethics," 2022, *arXiv:2206.11068*.
8. "Recommendation on the ethics of artificial intelligence," UNESCO, Paris, France SHS/BIO/PI/2021/1, 2022. [Online]. Available: <https://unesdoc.unesco.org/ark:/48223/pf0000381137>
9. P. de Laat, "Companies committed to responsible AI: From principles towards implementation and regulation?" *Philos. Technol.*, vol. 34, no. 4, pp. 1135–1193, 2021, doi: 10.1007/s13347-021-00474-3.
10. R. Tracy, "Biden administration developing National AI strategy," *Wall Street J.*, May 23, 2023. [Online]. Available: <https://www.wsj.com/articles/biden-administration-developing-national-ai-strategy-ee22b44e>
11. B. Mittelstadt, "Principles alone cannot guarantee ethical AI," *Nature Mach. Intell.*, vol. 1, no. 11, pp. 501–507, 2019, doi: 10.1038/s42256-019-0114-4.
12. R. Abraham, J. Schneider, and J. Brocke, "Data governance: A conceptual framework, structured review, and research agenda," *Int. J. Inf. Manage. Inf.*, vol. 49, pp. 424–438, Dec. 2019, doi: 10.1016/j.ijinfomgt.2019.07.008.
13. C. Bonina and B. Eaton, "Cultivating open government data platform ecosystems through governance: Lessons from Buenos Aires, Mexico City and Montevideo," *Government Inf. Quart.*, vol. 37, no. 3, 2020, Art. no. 101479, doi: 10.1016/j.giq.2020.101479.
14. World Bank, *World Development Report 2021: Data for Better Lives*. Washington, DC, USA: World Bank Publications, 2021.
15. V. Q. T. Li et al., "COVID-19, policy change, and post-pandemic data governance: A case analysis of contact tracing applications in East Asia," *Policy Soc.*, vol. 41, no. 1, pp. 129–142, Mar. 2022, doi: 10.1093/polsoc/puab019.
16. M. J. Parker, C. Fraser, L. Abeler-Dörner, and D. Bonsall, "Ethics of instantaneous contact tracing using mobile phone apps in the control of the COVID-19 pandemic," *J. Med. Ethics*, vol. 46, no. 7, pp. 427–431, 2020, doi: 10.1136/medethics-2020-106314.
17. S. Quach et al., "Digital technologies: Tensions in privacy and data," *J. Acad. Marketing Sci.*, vol. 50, no. 6, pp. 1299–1323, 2022, doi: 10.1007/s11747-022-00845-y.
18. M. Javaid, A. Haleem, S. Rab, R. Singh, and R. Suman, "Sensors for daily life: A review," *Sensors Int.*, vol. 2, Jul. 2021, Art. no. 100121, doi: 10.1016/j.sintl.2021.100121.
19. L. Martín, L. Sánchez, J. Lanza, and P. Sotres, "Development and evaluation of Artificial Intelligence techniques for IoT data quality assessment and curation," *Internet Things*, vol. 22, Jul. 2023, Art. no. 100779, doi: 10.1016/j.iot.2023.100779.
20. J. Park, M. Bernstein, R. Brewer, E. Kamar, and M. Morris, "Understanding the representation and representativeness of age in AI data sets," in *Proc. AAAI/ACM Conf. AI, Ethics, Soc. (AIES)*, New York, NY, USA: Association for Computing Machinery, 2021, pp. 834–842, doi: 10.1145/3461702.3462590.
21. Aapti Institute and Open Data Institute. "Enabling data sharing for social benefit through data trusts." GPAI. Accessed: Nov. 20, 2023. [Online]. Available: <https://gpai.ai/projects/data-governance/data-trusts/enabling-data-sharing-for-social-benefit-through-data-trusts.pdf>
22. D. S. Hain, R. Jurowetzki, M. Squicciarini, and L. Xu, "Unveiling the neurotechnology landscape: Scientific advancements innovations and major trends," UNESCO, Paris, France, 2023. [Online]. Available: <https://unesdoc.unesco.org/ark:/48223/pf0000386137>
23. U. Gasser and V. A. F. Almeida, "A layered model for AI governance," *IEEE Internet Comput.*, vol. 21,

- no. 6, pp. 58–62, Nov./Dec. 2017, doi: 10.1109/MIC.2017.4180835.
24. “Foundation models such as ChatGPT through the prism of the UNESCO recommendation on the ethics of artificial intelligence,” UNESCO, Paris, France, 2023. [Online]. Available: <https://unesdoc.unesco.org/ark:/48223/pf0000385629>
25. “ChatGPT.” OpenAI. Accessed: Nov. 20, 2023. [Online]. Available: <https://openai.com/chatgpt>
26. R. Grundke, L. Marcolin, T. Nguyen, and M. Squicciarini, “Which skills for the digital era? Returns to skills analysis,” Organisation for Economic Co-operation and Development, Paris, France, 2018. [Online]. Available: <https://www.oecd-ilibrary.org/>
27. L. Samek, M. Squicciarini, and E. Cammeraat, “The human capital behind AI: Jobs and skills demand from online job postings,” Organisation for Economic Co-operation and Development, Paris, France, 2018. [Online]. Available: [https://www.oecd-ilibrary.org/science-and-technology/the-human-capital-behind-ai\\_2e278150-en](https://www.oecd-ilibrary.org/science-and-technology/the-human-capital-behind-ai_2e278150-en)
28. E. Felten, M. Raj, and R. Seamans, “The occupational impact of artificial intelligence: Labor, skills, and polarization,” NYU Stern School of Business, New York, NY, USA, 2019. [Online]. Available: [https://papers.ssrn.com/sol3/papers.cfm?abstract\\_id=3368605](https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3368605)
29. M. L. Ameri, M. Schur, and D. Autor, “The labor market impact of technological change: From unbridled enthusiasm to qualified optimism to vast uncertainty,” National Bureau of Economic Research, Cambridge, MA, USA, NBER Working Paper 30074, 2022. [Online]. Available: <https://www.nber.org/papers/w30074>
30. L. Samek and S. Mariagrazia, “AI human capital, jobs and skills,” in *Impact of Artificial Intelligence in Business and Society: Opportunities and Challenges*, 1st ed. F. P. Appio, D. La Torre, F. Lazzeri, H. Masri, and F. Schiavone, Eds., Evanston, IL, USA: Routledge, 2023.
31. E. Strubell, A. Ganesh, and A. McCallum, “Energy and policy considerations for modern deep learning research,” in *Proc. AAAI Conf. Artif. Intell.*, 2020, vol. 34, no. 9, pp. 13,693–13,696, doi: 10.1609/aaai.v34i09.7123.
32. N. Maslej et al., “The AI index 2023 annual report,” AI Index Steering Committee, Institute for Human-Centered AI, Stanford Univ., Stanford, CA, USA, Apr. 2023. [Online]. Available: [https://aiindex.stanford.edu/wp-content/uploads/2023/04/HAI\\_AI-Index-Report\\_2023.pdf](https://aiindex.stanford.edu/wp-content/uploads/2023/04/HAI_AI-Index-Report_2023.pdf)
33. P. Li, J. Yang, M. A. Islam, and S. Ren, “Making AI less ‘Thirsty’: Uncovering and addressing the secret water footprint of AI models,” 2023. [Online]. Available: <https://arxiv.org/abs/2304.03271>
34. W. Gendron. “ChatGPT needs to ‘drink’ a water bottle’s worth of fresh water for every 20 to 50 questions you ask, researchers say.” Business Insider. Accessed: Nov. 20, 2023. [Online]. Available: <https://www.businessinsider.com/chatgpt-generative-ai-water-use-environmental-impact-study-2023-4>
35. A. L. Ligozat, J. Lefevre, A. Bugeau, and J. Combaz, “Unraveling the hidden environmental impacts of AI solutions for environment life cycle assessment of AI solutions,” *Sustainability*, vol. 14, no. 9, 2022, Art. no. 5172, doi: 10.3390/su14095172.
36. P. Dhar, “The carbon impact of artificial intelligence,” *Nature Mach. Intell.*, vol. 2, no. 8, pp. 423–425, 2020, doi: 10.1038/s42256-020-0219-9.
37. C. J. Wu et al., “Sustainable AI: Environmental implications, challenges and opportunities,” in *Proc. Mach. Learn. Syst.*, 2022, vol. 4, pp. 795–813.
38. A. Van Wynsberghe, “Sustainable AI: AI for sustainability and the sustainability of AI,” *AI Ethics*, vol. 1, no. 3, pp. 213–218, 2021, doi: 10.1007/s43681-021-00043-6.
39. “MEPs ready to negotiate first-ever rules for safe and transparent AI,” European Parliament, Strasbourg, France, Jun. 14, 2023. [Online]. Available: <https://www.europarl.europa.eu/news/en/press-room/20230609IPR96212/meps-ready-to-negotiate-first-ever-rules-for-safe-and-transparent-ai>
40. M. Barlow, *AI and Medicine: Data-Driven Strategies for Improving Healthcare and Saving Lives*. Sebastopol, CA, USA: O’Reilly Media, 2016.
41. “Staying ahead of the curve – The business case for responsible AI,” The Economist Intelligence Unit, London, U.K., 2020. [Online]. Available: <https://www.eiu.com/n/staying-ahead-of-the-curve-the-business-case-for-responsible-ai/>