United Nations
Educational, Scientific and
Cultural Organization

**Ad Hoc Expert Group (AHEG) for the preparation of a draft text of a recommendation on the ethics of artificial intelligence**

Distribution: limited

SHS/BIO/AHEG-AI/2020/4 REV.2
Paris, 7 September 2020
Original: English

# OUTCOME DOCUMENT:

# FIRST DRAFT OF THE RECOMMENDATION ON THE ETHICS OF ARTIFICIAL INTELLIGENCE

In line with the decision of UNESCO's General Conference at its 40th session (40 C/Resolution 37), the Director-General constituted the Ad Hoc Expert Group (AHEG) for the preparation of a draft text of a recommendation on the ethics of artificial intelligence in March 2020.

Adapting to the challenging situation posed by the COVID-19 pandemic, the AHEG worked virtually from the end of March until beginning of May 2020, and produced the first version of a draft text of the Recommendation on the Ethics of Artificial Intelligence.

An extensive multi-stakeholder consultation process on this first version was conducted from June to August 2020, based on three components: (i) public online consultation (receiving more than 800 responses); (ii) regional and subregional virtual consultations co-organized with host countries/institutions in all of UNESCO's regions (involving more than 500 participants); and (iii) open, multi-stakeholder, and citizen deliberation workshops organized by partners (involving approximately 500 participants). The consultation process generated more than 50,000 comments on the text.

Taking into account the feedback received during this consultation process, the AHEG revised the first version of the draft text from August until beginning of September 2020 to produce the first draft of the Recommendation contained in this document, which will be transmitted to Member States for written comments in September 2020.

The AHEG was supported by the Assistant Director-General for Social and Human Sciences, and the Bioethics and Ethics of Science Section.

This document does not claim to be exhaustive and does not necessarily represent the views of the Member States of UNESCO.

# FIRST DRAFT OF THE RECOMMENDATION ON THE ETHICS OF ARTIFICIAL INTELLIGENCE

## PREAMBLE

The General Conference of the United Nations Educational, Scientific and Cultural Organization (UNESCO), meeting in Paris from xx to xx, at its xx session,

***Recognizing*** the profound and dynamic impact of artificial intelligence (AI) on societies, ecosystems, and human lives, including the human mind, in part because of the new ways in which it influences human thinking, interaction and decision-making, and affects education, human, social and natural sciences, culture, and communication and information,

***Recalling*** that, by the terms of its Constitution, UNESCO seeks to contribute to peace and security by promoting collaboration among nations through education, the sciences, culture, and communication and information, in order to further universal respect for justice, for the rule of law and for the human rights and fundamental freedoms which are affirmed for the peoples of the world,

***Convinced*** that the standard-setting instrument presented here, based on international law and on a global normative approach, focusing on human dignity and human rights, as well as gender equality, social and economic justice, physical and mental well-being, diversity, interconnectedness, inclusiveness, and environmental and ecosystem protection can guide AI technologies in a responsible direction,

***Considering*** that AI technologies can be of great service to humanity but also raise fundamental ethical concerns, for instance regarding the biases they can embed and exacerbate, potentially resulting in inequality, exclusion and a threat to cultural, social and ecological diversity and social or economic divides; the need for transparency and understandability of the workings of algorithms and the data with which they have been trained; and their potential impact on human dignity, human rights, gender equality, privacy, freedom of expression, access to information, social, economic, political and cultural processes, scientific and engineering practices, animal welfare, and the environment and ecosystems,

***Recognizing*** that AI technologies can deepen existing divides and inequalities in the world, within and between countries, and that justice, trust and fairness must be upheld so that no one should be left behind, either in enjoying the benefits of AI technologies or in the protection against their negative implications, while recognizing the different circumstances of different countries and the desire of some people not to take part in all technological developments,

***Conscious*** of the fact that all countries are facing an acceleration of the use of information and communication technologies and AI technologies, as well as an increasing need for media and information literacy, and that the digital economy presents important societal, economic and environmental challenges and opportunities of benefits sharing, especially for low- and middle-income countries (LMICs), including but not limited to least developed countries (LDCs), landlocked developing countries (LLDCs) and small island developing States (SIDS), requiring the recognition, protection and promotion of endogenous cultures, values and knowledge in order to develop sustainable digital economies,

***Recognizing*** that AI technologies have the potential to be beneficial to the environment and ecosystems but in order for those benefits to be realized, fair access to the technologies is required without ignoring but instead addressing potential harms to and impact on the environment and ecosystems,

***Noting*** that addressing risks and ethical concerns should not hamper innovation but rather provide new opportunities and stimulate new and responsible practices of research and innovation that anchor AI technologies in human rights, values and principles, and moral and ethical reflection,

***Recalling*** that in November 2019, the General Conference of UNESCO, at its 40th session, adopted 40 C/Resolution 37, by which it mandated the Director-General "to prepare an international standard-setting instrument on the ethics of artificial intelligence (AI) in the form of a recommendation", which is to be submitted to the General Conference at its 41st session in 2021,

***Recognizing*** that the development of AI technologies results in an increase of information which necessitates a commensurate increase in media and information literacy as well as access to critical sources of information,

***Observing*** that a normative framework for AI technologies and its social implications finds its basis in ethics, as well as human rights, fundamental freedoms, access to data, information and knowledge, international and national legal frameworks, the freedom of research and innovation, human and environmental and ecosystem well-being, and connects ethical values and principles to the challenges and opportunities linked to AI technologies, based on common understanding and shared aims,

***Recognizing*** that ethical values and principles can powerfully shape the development and implementation of rights-based policy measures and legal norms, by providing guidance where the ambit of norms is unclear or where such norms are not yet in place due to the fast pace of technological development combined with the relatively slower pace of policy responses,

***Convinced*** that globally accepted ethical standards for AI technologies and international law, in particular human rights law, principles and standards can play a key role in harmonizing AI-related legal norms across the globe,

***Recognizing*** the Universal Declaration of Human Rights (1948), including Article 27 emphasizing the right to share in scientific advancement and its benefits; the instruments of the international human rights framework, including the International Convention on the Elimination of All Forms of Racial Discrimination (1965), the International Covenant on Civil and Political Rights (1966), the International Covenant on Economic, Social and Cultural Rights (1966), the United Nations Convention on the Elimination of All Forms of Discrimination against Women (1979), the United Nations Convention on the Rights of the Child (1989), and the United Nations Convention on the Rights of Persons with Disabilities (2006); the UNESCO Convention on the Protection and Promotion of the Diversity of Cultural Expressions (2005),

***Noting*** the UNESCO Declaration on the Responsibilities of the Present Generations Towards Future Generations (1997); the United Nations Declaration on the Rights of Indigenous Peoples (2007); the Report of the United Nations Secretary-General on the Follow-up to the Second World Assembly on Ageing (A/66/173) of 2011, focusing on the situation of the human rights of older persons; the Report of the Special Representative of the United Nations Secretary-General on the issue of human rights and transnational corporations and other business enterprises (A/HRC/17/31) of 2011, outlining the "Guiding Principles on Business and Human Rights: Implementing United Nations 'Protect, Respect and Remedy' Framework"; the United Nations General Assembly resolution on the review of the World Summit on the Information Society (A/68/302); the Human Rights Council's resolution on "The right to privacy in the digital age" (A/HRC/RES/42/15) adopted on 26 September 2019; the Report of the Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression (A/73/348); the UNESCO Recommendation on Science and Scientific Researchers (2017); the UNESCO Internet Universality Indicators (endorsed by UNESCO's International Programme for the Development of Communication in 2019), including the R.O.A.M. principles (endorsed by UNESCO's General Conference in 2015); the UNESCO Recommendation Concerning the Preservation of, and Access to, Documentary Heritage Including in Digital Form (2015); the Report of the United Nations Secretary-General's High-level Panel on Digital Cooperation on "The Age of Digital Interdependence" (2019), and the United Nations Secretary-General's Roadmap for Digital Cooperation (2020); the Universal Declaration on Bioethics and Human Rights (2005); the UNESCO Declaration on Ethical Principles in relation to Climate Change (2017); the United Nations Global Pulse initiative; and the outcomes and reports of the ITU's AI for Good Global Summits,

***Noting also*** existing frameworks related to the ethics of AI of other intergovernmental organizations, such as the relevant human rights and other legal instruments adopted by the Council of Europe, and the work of its Ad Hoc Committee on AI (CAHAI); the work of the European Union related to AI, and of the European Commission's High-Level Expert Group on AI, including the Ethical Guidelines for Trustworthy AI; the work of OECD's first Group of Experts (AIGO) and its successor the OECD Network of Experts on AI (ONE AI), the OECD's Recommendation of the Council on AI and the OECD AI Policy Observatory (OECD.AI); the G20 AI Principles, drawn therefrom, and outlined in the G20 Ministerial Statement on Trade and Digital Economy; the G7's Charlevoix Common Vision for the Future of AI; the work of the African Union's Working Group on AI; and the work of the Arab League's Working Group on AI,

***Emphasizing*** that specific attention must be paid to LMICs, including but not limited to LDCs, LLDCs and SIDS, as they have their own capacity but have been underrepresented in the AI ethics debate, which raises concerns about neglecting local knowledge, cultural and ethical pluralism, value systems and the demands of global fairness to deal with the positive and negative impacts of AI technologies,

***Conscious*** of the many existing national policies and other frameworks related to the ethics and regulation of AI technologies,

***Conscious as well*** of the many initiatives and frameworks related to the ethics of AI developed by the private sector, professional organizations, and non-governmental organizations, such as the IEEE's Global Initiative on Ethics of Autonomous and Intelligent Systems and its work on Ethically Aligned Design; the World Economic Forum's "Global Technology Governance: A Multistakeholder Approach"; the UNI Global Union's "Top 10 Principles for Ethical Artificial Intelligence"; the Montreal Declaration for a Responsible Development of AI; the Toronto Declaration: Protecting the rights to equality and non-discrimination in machine learning systems; the Harmonious Artificial Intelligence Principles (HAIP); and the Tenets of the Partnership on AI,

***Convinced*** that AI technologies can bring important benefits, but that achieving them can also amplify tension around innovation debt, asymmetric access to knowledge, barriers of rights to information and gaps in capacity of creativity in developing cycles, human and institutional capacities, barriers to access to technological innovation, and a lack of adequate physical and digital infrastructure and regulatory frameworks regarding data,

***Underlining*** that global cooperation and solidarity are needed to address the challenges that AI technologies bring in diversity and interconnectivity of cultures and ethical systems, to mitigate potential misuse, and to ensure that AI strategies and regulatory frameworks are not guided only by national and commercial interests and economic competition,

***Taking fully into account*** that the rapid development of AI technologies challenges their ethical implementation and governance, because of the diversity of ethical orientations and cultures around the world, the lack of agility of the law in relation to technology and knowledge societies, and the risk that local and regional ethical standards and values be disrupted by AI technologies,

1. ***Adopts*** the present Recommendation on the Ethics of Artificial Intelligence;

2. ***Recommends*** that Member States apply the provisions of this Recommendation by taking appropriate steps, including whatever legislative or other measures may be required, in conformity with the constitutional practice and governing structures of each State, to give effect within their jurisdictions to the principles and norms of the Recommendation in conformity with international law, as well as constitutional practice;

3. ***Also recommends*** that Member States ensure assumption of responsibilities by all stakeholders, including private sector companies in AI technologies, and bring the Recommendation to the attention of the authorities, bodies, research and academic organizations, institutions and

organizations in public, private and civil society sectors involved in AI technologies, in order to guarantee that the development and use of AI technologies are guided by both sound scientific research as well as ethical analysis and evaluation.

## I. SCOPE OF APPLICATION

1. This Recommendation addresses ethical issues related to AI. It approaches AI ethics as a systematic normative reflection, based on a holistic and evolving framework of interdependent values, principles and actions that can guide societies in dealing responsibly with the known and unknown impacts of AI technologies on human beings, societies, and the environment and ecosystems, and offers them a basis to accept or reject AI technologies. Rather than equating ethics to law, human rights, or a normative add-on to technologies, it considers ethics as a dynamic basis for the normative evaluation and guidance of AI technologies, referring to human dignity, well-being and the prevention of harm as a compass and rooted in the ethics of science and technology.

2. This Recommendation does not have the ambition to provide one single definition of AI, since such a definition would need to change over time, in accordance with technological developments. Rather, its ambition is to address those features of AI systems that are of central ethical relevance and on which there is large international consensus. Therefore, this Recommendation approaches AI systems as technological systems which have the capacity to process information in a way that resembles intelligent behaviour, and typically includes aspects of reasoning, learning, perception, prediction, planning or control. Three elements have a central place in this approach:

(a) AI systems are information-processing technologies that embody models and algorithms that produce a capacity to learn and to perform cognitive tasks leading to outcomes such as prediction and decision-making in real and virtual environments. AI systems are designed to operate with some aspects of autonomy by means of knowledge modelling and representation and by exploiting data and calculating correlations. AI systems may include several methods, such as but not limited to:

(i) machine learning, including deep learning and reinforcement learning,

(ii) machine reasoning, including planning, scheduling, knowledge representation and reasoning, search, and optimization, and

(iii) cyber-physical systems, including the Internet-of-Things, robotic systems, social robotics, and human-computer interfaces which involve control, perception, the processing of data collected by sensors, and the operation of actuators in the environment in which AI systems work.

(b) Ethical questions regarding AI systems pertain to all stages of the AI system life cycle, understood here to range from research, design, and development to deployment and use, including maintenance, operation, trade, financing, monitoring and evaluation, validation, end-of-use, disassembly, and termination. In addition, AI actors can be defined as any actor involved in at least one stage of the AI life cycle, and can refer both to natural and legal persons, such as researchers, programmers, engineers, data scientists, end-users, large technology companies, small and medium enterprises, start-ups, universities, public entities, among others.

(c) AI systems raise new types of ethical issues that include, but are not limited to, their impact on decision-making, employment and labour, social interaction, health care, education, media, freedom of expression, access to information, privacy, democracy, discrimination, and weaponization. Furthermore, new ethical challenges are created by the potential of AI algorithms to reproduce biases, for instance regarding gender, ethnicity, and age, and thus to exacerbate already existing forms of discrimination, identity prejudice and stereotyping. Some of these issues are related to the capacity of

AI systems to perform tasks which previously only living beings could do, and which were in some cases even limited to human beings only. These characteristics give AI systems a profound, new role in human practices and society, as well as in their relationship with the environment and ecosystems, creating a new context for children and young people to grow up in, develop an understanding of the world and themselves, critically understand media and information, and learn to make decisions. In the long term, AI systems could challenge human's special sense of experience and agency, raising additional concerns about human self-understanding, social, cultural and environmental interaction, autonomy, agency, worth and dignity.

3. This Recommendation pays specific attention to the broader ethical implications of AI systems in relation to the central domains of UNESCO: education, science, culture, and communication and information, as explored in the 2019 Preliminary Study on the Ethics of Artificial Intelligence by the UNESCO World Commission on Ethics of Scientific Knowledge and Technology (COMEST):

(a) Education, because living in digitalizing societies requires new educational practices, the need for ethical reflection, critical thinking, responsible design practices, and new skills, given the implications for the labour market and employability.

(b) Science, in the broadest sense and including all academic fields from the natural sciences and medical sciences to the social sciences and humanities, as AI technologies bring new research capacities, have implications for our concepts of scientific understanding and explanation, and create a new basis for decision-making.

(c) Cultural identity and diversity, as AI technologies can enrich cultural and creative industries, but can also lead to an increased concentration of supply of cultural content, data, markets, and income in the hands of only a few actors, with potential negative implications for the diversity and pluralism of languages, media, cultural expressions, participation and equality.

(d) Communication and information, as AI technologies play an increasingly important role in the processing, structuring and provision of information, and the issues of automated journalism and the algorithmic provision of news and moderation and curation of content on social media and search engines are just a few examples raising issues related to access to information, disinformation, misinformation, misunderstanding, the emergence of new forms of societal narratives, discrimination, freedom of expression, privacy, and media and information literacy, among others.

4. This Recommendation is addressed to States, both as AI actors and as responsible for developing legal and regulatory frameworks throughout the entire AI system life cycle, and for promoting business responsibility. It also provides ethical guidance to all AI actors, including the private sector, by providing a basis for an Ethical Impact Assessment of AI systems throughout their life cycle.

## II. AIMS AND OBJECTIVES

5. This 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.

6. In addition to the ethical frameworks regarding AI that have already been developed by various organizations all over the world, this Recommendation aims to bring a globally accepted normative instrument that does not only focus on the articulation of values and principles, but also on their practical realization, via concrete policy recommendations, with a strong emphasis on issues of gender equality and protection of the environment and ecosystems.

7. Because the complexity of the ethical issues surrounding AI necessitates the cooperation of multiple stakeholders across the various levels and sectors of international, regional and national communities, this Recommendation aims to enable stakeholders to take shared responsibility based on a global and intercultural dialogue.

8. The objectives of this Recommendation are:

(a) to provide a universal framework of values, principles and actions to guide States in the formulation of their legislation, policies or other instruments regarding AI;

(b) to guide the actions of individuals, groups, communities, institutions and private sector companies to ensure the embedding of ethics in all stages of the AI system life cycle;

(c) to promote respect for human dignity and gender equality, to safeguard the interests of present and future generations, and to protect human rights, fundamental freedoms, and the environment and ecosystems in all stages of the AI system life cycle;

(d) to foster multi-stakeholder, multidisciplinary and pluralistic dialogue about ethical issues relating to AI systems; and

(e) to promote equitable access to developments and knowledge in the field of AI and the sharing of benefits, with particular attention to the needs and contributions of LMICs, including LDCs, LLDCs and SIDS.

## III. VALUES AND PRINCIPLES

9. The values and principles included below should be respected by all actors in the AI system life cycle, in the first place, and be promoted through amendments to existing and elaboration of new legislation, regulations and business guidelines. This must comply with international law as well as with international human rights law, principles and standards, and should be in line with social, political, environmental, educational, scientific and economic sustainability objectives.

10. Values play a powerful role as motivating ideals in shaping policy measures and legal norms. While the set of values outlined below thus inspires desirable behaviour and represents the foundations of principles, the principles unpack the values underlying them more concretely so that the values can be more easily operationalized in policy statements and actions.

11. While all the values and principles outlined below are desirable per se, in any practical context there are inevitable trade-offs among them, requiring complex choices to be made about contextual prioritization, without compromising other principles or values in the process. Trade-offs should take account of concerns related to proportionality and legitimate purpose. To navigate such scenarios judiciously will typically require engagement with a broad range of appropriate stakeholders guided by international human rights law, standards and principles, making use of social dialogue, as well as ethical deliberation, due diligence, and impact assessment.

12. The trustworthiness and integrity of the life cycle of AI systems, if achieved, work for the good of humanity, individuals, societies, and the environment and ecosystems, and embody the values and principles set out in this Recommendation. People should have good reason to trust that AI systems bring shared benefits, while adequate measures are taken to mitigate risks. An essential requirement for trustworthiness is that, throughout their life cycle, AI systems are subject to monitoring by governments, private sector companies, independent civil society and other stakeholders. As trustworthiness is an outcome of the operationalization of the principles in this document, the policy actions proposed in this Recommendation are all directed at promoting trustworthiness in all stages of the AI life cycle.

## III.1. VALUES

### Respect, protection and promotion of human dignity, human rights and fundamental freedoms

13. The dignity of every human person constitutes a foundation for the indivisible system of human rights and fundamental freedoms and is essential throughout the life cycle of AI systems. Human dignity relates to the recognition of the intrinsic worth of each individual human being and thus dignity is not tied to sex, gender, language, religion, political or other opinion, national, ethnic, indigenous or social origin, sexual orientation and gender identity, property, birth, disability, age or other status.

14. No human being should be harmed physically, economically, socially, politically, or mentally during any phase of the life cycle of AI systems. Throughout the life cycle of AI systems the quality of life of every human being should be enhanced, while the definition of “quality of life” should be left open to individuals or groups, as long as there is no violation or abuse of human rights, or the dignity of humans in terms of this definition.

15. Persons may interact with AI systems throughout their life cycle and receive assistance from them such as care for vulnerable people, including but not limited to children, older persons, persons with disabilities or the ill. Within such interactions, persons should never be objectified, nor should their dignity be undermined, or human rights violated or abused.

16. Human rights and fundamental freedoms must be respected, protected, and promoted throughout the life cycle of AI systems. Governments, private sector, civil society, international organizations, technical communities, and academia must respect human rights instruments and frameworks in their interventions in the processes surrounding the life cycle of AI systems. New technologies need to provide new means to advocate, defend and exercise human rights and not to infringe them.

### Environment and ecosystem flourishing

17. Environmental and ecosystem flourishing should be recognized and promoted through the life cycle of AI systems. Furthermore, environment and ecosystems are the existential necessity for humanity and other living beings to be able to enjoy the benefits of advances in AI.

18. All actors involved in the life cycle of AI systems must follow relevant international law and domestic legislation, standards and practices, such as precaution, designed for environmental and ecosystem protection and restoration, and sustainable development. They should reduce the environmental impact of AI systems, including but not limited to, its carbon footprint, to ensure the minimization of climate change and environmental risk factors, and prevent the unsustainable exploitation, use and transformation of natural resources contributing to the deterioration of the environment and the degradation of ecosystems.

### Ensuring diversity and inclusiveness

19. Respect, protection and promotion of diversity and inclusiveness should be ensured throughout the life cycle of AI systems, at a minimum consistent with international human rights law, standards and principles, as well as demographic, cultural, gender and social diversity and inclusiveness. This may be done by promoting active participation of all individuals or groups based on sex, gender, language, religion, political or other opinion, national, ethnic, indigenous or social origin, sexual orientation and gender identity, property, birth, disability, age or other status, in the life cycle of AI systems. Any homogenizing tendency should be monitored and addressed.

20. The scope of lifestyle choices, beliefs, opinions, expressions or personal experiences, including the optional use of AI systems and the co-design of these architectures should not be restricted in any way during any phase of the life cycle of AI systems.

21. Furthermore, efforts should be made to overcome, and never exploit, the lack of necessary technological infrastructure, education and skills, as well as legal frameworks, in some communities, and particularly in LMICs, LDCs, LLDCs and SIDS.

**Living in harmony and peace**

22. AI actors should play an enabling role for harmonious and peaceful life, which is to ensure an interconnected future ensuring the benefit of all. The value of living in harmony and peace points to the potential of AI systems to contribute throughout their life cycle to the interconnectedness of all living creatures with each other and with the natural environment.

23. The notion of humans being interconnected is based on the knowledge that every human belongs to a greater whole, which is diminished when others are diminished in any way. Living in harmony and peace requires an organic, immediate, uncalculated bond of solidarity, characterized by a permanent search for non-conflictual, peaceful relations, tending towards consensus with others and harmony with the natural environment in the broadest sense of the term.

24. This value demands that peace should be promoted throughout the life cycle of AI systems, in so far as the processes of the life cycle of AI systems should not segregate, objectify, or undermine the safety of human beings, divide and turn individuals and groups against each other, or threaten the harmonious coexistence between humans, non-humans, and the natural environment, as this would negatively impact on humankind as a collective.

## III.2. PRINCIPLES

**Proportionality and do no harm**

25. It should be recognized that AI technologies do not necessarily, per se, ensure human and environmental and ecosystem flourishing. Furthermore, none of the processes related to the AI system life cycle shall exceed what is necessary to achieve legitimate aims or objectives and should be appropriate to the context. In the event of possible occurrence of any harm to human beings 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.

26. The choice of an AI method should be justified in the following ways: (a) The AI method chosen should be desirable and proportional to achieve a given legitimate aim; (b) The AI method chosen should not have a negative infringement on the foundational values captured in this document; (c) The AI method should be appropriate to the context and should be based on rigorous scientific foundations. In scenarios that involve life and death decisions, final human determination should apply.

**Safety and security**

27. Unwanted harms (safety risks) and vulnerabilities to attacks (security risks) should be avoided throughout the life cycle of AI systems to ensure human and environmental and ecosystem safety and security. Safe and secure AI will be enabled by the development of sustainable, privacy-protective data access frameworks that foster better training of AI models utilizing quality data.

**Fairness and non-discrimination**

28. AI actors should promote social justice, by respecting fairness. Fairness implies sharing benefits of AI technologies at local, national and international levels, while taking into consideration the specific needs of different age groups, cultural systems, different language groups, persons with disabilities, girls and women, and disadvantaged, marginalized and vulnerable populations. At the local level, it is a matter of working to give communities access to AI systems in the languages of their choice and respecting different cultures. At the national level, governments are obliged to demonstrate equity between rural and urban areas, and among all persons without distinction as to

sex, gender, language, religion, political or other opinion, national, ethnic, indigenous or social origin, sexual orientation and gender identity, property, birth, disability, age or other status, in terms of access to and participation in the AI system life cycle. At the international level, the most technologically advanced countries have an obligation of solidarity with the least advanced to ensure that the benefits of AI technologies are shared such that access to and participation in the AI system life cycle for the latter contributes to a fairer world order with regard to information, communication, culture, education, research, and socio-economic and political stability.

29. AI actors should make all efforts to minimize and avoid reinforcing or perpetuating inappropriate socio-technical biases based on identity prejudice, throughout the life cycle of the AI system to ensure fairness of such systems. There should be a possibility to have a remedy against unfair algorithmic determination and discrimination.

30. Furthermore, discrimination, digital and knowledge divides, and global inequalities need to be addressed throughout an AI system life cycle, including in terms of access to technology, data, connectivity, knowledge and skills, and participation of the affected communities as part of the design phase, such that every person is treated equitably.

**Sustainability**

31. The development of sustainable societies relies on the achievement of a complex set of objectives on a continuum of social, cultural, economic and environmental dimensions. The advent of AI technologies can either benefit sustainability objectives or hinder their realization, depending on how they are applied across countries with varying levels of development. The continuous assessment of the social, cultural, economic and environmental impact of AI technologies should therefore be carried out with full cognizance of the implications of AI technologies for sustainability as a set of constantly evolving goals across a range of dimensions, such as currently identified in the United Nations Sustainable Development Goals (SDGs).

**Privacy**

32. 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 both at the personal and collective level. It is crucial that data for AI is being collected, used, shared, archived and deleted in ways that are consistent with the values and principles set forth in this Recommendation.

33. Adequate data protection frameworks and governance mechanisms should be established by regulatory agencies, at national or supranational level, protected by judicial systems, and ensured throughout the life cycle of AI systems. This protection framework and mechanisms concern the collection, control over, and use of data and exercise of their rights by data subjects and of the right for individuals to have personal data erased, ensuring a legitimate aim and a valid legal basis for the processing of personal data as well as for the personalization, and de- and re-personalization of data, transparency, appropriate safeguards for sensitive data, and effective independent oversight.

34. Algorithmic systems require thorough privacy impact assessments which also include societal and ethical considerations of their use and an innovative use of the privacy by design approach.

**Human oversight and determination**

35. It must always be possible to attribute ethical and legal responsibility for any stage of the life cycle of AI systems to physical persons or to existing legal entities. Human oversight refers thus not only to individual human oversight, but to public oversight, as appropriate.

36. It may be the case that sometimes humans would have to rely on AI systems for reasons of efficacy, but the decision to cede control in limited contexts remains that of humans, as humans can

resort to AI systems in decision-making and acting, but an AI system can never replace ultimate human responsibility and accountability.

**Transparency and explainability**

37. The transparency of AI systems is often a crucial precondition to ensure that fundamental human rights and ethical principles are respected, protected and promoted. Transparency is necessary for relevant national and international liability legislation to work effectively.

38. While efforts need to be made to increase transparency and explainability of AI systems throughout their life cycle to support democratic governance, the level of transparency and explainability should always be appropriate to the context, as some trade-offs exist between transparency and explainability and other principles such as safety and security. People have the right to be aware when a decision is being made on the basis of AI algorithms, and in those circumstances require or request explanatory information from private sector companies or public sector institutions.

39. From a socio-technical lens, greater transparency contributes to more peaceful, just and inclusive societies. It allows for public scrutiny that can decrease corruption and discrimination, and can also help detect and prevent negative impacts on human rights. Transparency may contribute to trust from humans for AI systems. Specific to the AI system, transparency can enable people to understand how each stage of an AI system is put in place, appropriate to the context and sensitivity of the AI system. It may also include insight into factors that impact a specific prediction or decision, and whether or not appropriate assurances (such as safety or fairness measures) are in place. In cases where serious adverse human rights impacts are foreseen, transparency may also require the sharing of specific code or datasets.

40. Explainability refers to making intelligible and providing insight into the outcome of AI systems. The explainability of AI systems also refers to the understandability of the input, output and behaviour of each algorithmic building block and how it contributes to the outcome of the systems. Thus, explainability is closely related to transparency, as outcomes and sub-processes leading to outcomes should be understandable and traceable, appropriate to the use context.

41. Transparency and explainability relate closely to adequate responsibility and accountability measures, as well as to the trustworthiness of AI systems.

**Responsibility and accountability**

42. AI actors should respect, protect and promote human rights and promote the protection of the environment and ecosystems, assuming ethical and legal responsibility in accordance with extant national and international law, in particular international human rights law, principles and standards, and ethical guidance throughout the life cycle of AI systems. The ethical responsibility and liability for the decisions and actions based in any way on an AI system should always ultimately be attributable to AI actors.

43. Appropriate oversight, impact assessment, 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 (the working of) AI systems in particular to address any conflicts with human rights and threats to environmental and ecosystem well-being.

**Awareness and literacy**

44. Public awareness and understanding of AI technologies and the value of data should be promoted through open and accessible education, civic engagement, digital skills and AI ethics training, media and information literacy and training led jointly by governments, intergovernmental organizations, civil society, academia, the media, community leaders and the private sector, and

considering the existing linguistic, social and cultural diversity, to ensure effective public participation so that all members of society can take informed decisions about their use of AI systems and be protected from undue influence.

45. Learning about the impact of AI systems should include learning about, through and for human rights, meaning that the approach and understanding of AI systems should be grounded by their impact on human rights and access to rights.

**Multi-stakeholder and adaptive governance and collaboration**

46. International law and sovereignty should be respected in the use of data. Data sovereignty means that States, complying with international law, regulate the data generated within or passing through their territories, and take measures towards effective regulation of data based on respect for the right to privacy and other human rights.

47. Participation of different stakeholders throughout the AI system life cycle is necessary for inclusive AI governance, sharing of benefits of AI, and fair technological advancement and its contribution to development goals. Stakeholders include but are not limited to governments, intergovernmental organizations, the technical community, civil society, researchers and academia, media, education, policy-makers, private sector companies, human rights institutions and equality bodies, anti-discrimination monitoring bodies, and groups for youth and children. The adoption of open standards and interoperability to facilitate collaboration must be in place. Measures must be adopted to take into account shifts in technologies, the emergence of new groups of stakeholders, and to allow for meaningful intervention by marginalized groups, communities and individuals.

## IV. AREAS OF POLICY ACTION

48. The policy actions described in the following policy areas operationalize the values and principles set out in this Recommendation. The main action is for Member States to put in place policy frameworks or mechanisms and to ensure that other stakeholders, such as private sector companies, academic and research institutions, and civil society, adhere to them by, among other actions, assisting all stakeholders to develop ethical impact assessment and due diligence tools. The process for developing such policies or mechanisms should be inclusive of all stakeholders and should take into account the circumstances and priorities of each Member State. UNESCO can be a partner and support Member States in the development as well as monitoring and evaluation of policy mechanisms.

49. UNESCO recognizes that Member States will be at different stages of readiness to implement this Recommendation, in terms of scientific, technological, economic, educational, legal, regulatory, infrastructural, societal, cultural and other dimensions. It is noted that “readiness” here is a dynamic status. In order to enable the effective implementation of this Recommendation, UNESCO will therefore: (1) develop a readiness assessment methodology to assist Member States in identifying their status at specific moments of their readiness trajectory along a continuum of dimensions; and (2) ensure support for Member States in terms of developing a globally accepted methodology for Ethical Impact Assessment (EIA) of AI technologies, sharing of best practices, assessment guidelines and other mechanisms and analytical work.

### POLICY AREA 1: ETHICAL IMPACT ASSESSMENT

50. Member States should introduce impact assessments to identify and assess benefits, concerns and risks of AI systems, as well as risk prevention, mitigation and monitoring measures. The ethical impact assessment should identify impacts on human rights, in particular but not limited to the rights of vulnerable groups, labour rights, the environment and ecosystems, and ethical and social implications in line with the principles set forth in this Recommendation.

51. 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 human rights, rule of law and inclusive societies. Member States should also be able to assess the socio-economic impact of AI systems on poverty and ensure that the gap between people living in wealth and poverty, as well as the digital divide among and within countries are not increased with the massive adoption of AI technologies at present and in the future. In order to do this, enforceable transparency protocols should be implemented, corresponding to the right of access to information, including information of public interest held by private entities.

52. Member States and private sector companies should implement proper measures to monitor all phases of an AI system life cycle, including the behaviour 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.

53. Governments should adopt a regulatory framework that sets out a procedure, particularly for public authorities, to carry out ethical impact assessments on AI systems to predict consequences, mitigate risks, avoid harmful consequences, facilitate citizen participation and address societal challenges. The assessment should also establish appropriate oversight mechanisms, including auditability, traceability and explainability which enable the assessment of algorithms, data and design processes, as well as include external review of AI systems. Ethical impact assessments carried out by public authorities should be transparent and open to the public. Such assessments should also be multidisciplinary, multi-stakeholder, multicultural, pluralistic and inclusive. Member States are encouraged to put in place mechanisms and tools, for example regulatory sandboxes or testing centres, which would enable impact monitoring and assessment in a multidisciplinary and multi-stakeholder fashion. The public authorities should be required to monitor the AI systems implemented and/or deployed by those authorities by introducing appropriate mechanisms and tools.

54. Member States should establish monitoring and evaluation mechanisms for initiatives and policies related to AI ethics. Possible mechanisms include: a repository covering human rights-compliant and ethical development of AI systems; a lessons sharing mechanism for Member States to seek feedback from other Member States on their policies and initiatives; a guide for all AI actors to assess their adherence to policy recommendations mentioned in this document; and follow-up tools. International human rights law, standards and principles should form part of the ethical aspects of AI system assessments.

**POLICY AREA 2: ETHICAL GOVERNANCE AND STEWARDSHIP**

55. Member States should ensure that any AI governance mechanism is inclusive, transparent, multidisciplinary, multilateral (this includes the possibility of mitigation and redress of harm across borders), and multi-stakeholder. Governance should include aspects of anticipation, protection, monitoring of impact, enforcement and redressal.

56. Member States should ensure that harms caused to users through AI systems are investigated and redressed, by enacting strong enforcement mechanisms and remedial actions, to make certain that human rights and the rule of law are respected in the digital world as it is in the physical world. Such mechanisms and actions should include remediation mechanisms provided by private sector companies. The auditability and traceability of AI systems should be promoted to this end. In addition, Member States should strengthen their institutional capacities to deliver on this duty of care and should collaborate with researchers and other stakeholders to investigate, prevent and mitigate any potentially malicious uses of AI systems.

57. Member States are encouraged to consider forms of soft governance such as a certification mechanism for AI systems and the mutual recognition of their certification, according to the sensitivity of the application domain and expected impact on human lives, the environment and ecosystems, and other ethical considerations set forth in this Recommendation. Such a mechanism might include different levels of audit of systems, data, and adherence to ethical guidelines, and should be

validated by authorized parties in each country. At the same time, such a mechanism must not hinder innovation or disadvantage small and medium enterprises or start-ups by requiring large amounts of paperwork. These mechanisms would also include a regular monitoring component to ensure system robustness and continued integrity and adherence to ethical guidelines over the entire life cycle of the AI system, requiring re-certification if necessary.

58. Government and public authorities should be required to carry out self-assessment of existing and proposed AI systems, which in particular, should include the assessment whether the adoption of AI is appropriate and, if so, should include further assessment to determine what the appropriate method is, as well as assessment as to whether such adoption transgresses any human rights law, standards and principles.

59. Member States should encourage public entities, private sector companies and civil society organizations to involve different stakeholders in their AI governance and to consider adding the role of an independent AI Ethics Officer or some other mechanism to oversee ethical impact assessment, auditing and continuous monitoring efforts and ensure ethical guidance of AI systems. Member States, private sector companies and civil society organizations, with the support of UNESCO, are encouraged to create a network of independent AI Ethics Officers to give support to this process at national, regional and international levels.

60. Member States should foster the development of, and access to, a digital ecosystem for ethical development of AI systems at the national level, while encouraging international collaboration. Such an ecosystem includes in particular digital technologies and infrastructure, and mechanisms for sharing AI knowledge, as appropriate. In this regard, Member States should consider reviewing their policies and regulatory frameworks, including on access to information and open government to reflect AI-specific requirements and promoting mechanisms, such as open repositories for publicly-funded or publicly-held data and source code and data trusts, to support the safe, fair, legal and ethical sharing of data, among others.

61. Member States should establish mechanisms, in collaboration with international organizations, transnational corporations, academic institutions and civil society, to ensure the active participation of all Member States, especially LMICs, in particular LDCs, LLDCs and SIDS, in international discussions concerning AI governance. This can be through the provision of funds, ensuring equal regional participation, or any other mechanisms. Furthermore, in order to ensure the inclusiveness of AI fora, Member States should facilitate the travel of AI actors in and out of their territory, especially from LMICs, in particular LDCs, LLDCs and SIDS, for the purpose of participating in these fora.

62. Amendments to existing or elaboration of new national legislation addressing AI systems must comply with international human rights law and promote human rights and fundamental freedoms throughout the AI system life cycle. Promotion thereof should also take the form of governance initiatives, good exemplars of collaborative practices regarding AI systems, and national and international technical and methodological guidelines as AI technologies advance. Diverse sectors, including the private sector, in their practices regarding AI systems must respect, protect and promote human rights and fundamental freedoms using existing and new instruments in combination with this Recommendation.

63. Member States should provide mechanisms for human rights and for social and economic impact of AI monitoring and oversight, and other governance mechanisms such as independent data protection authorities, sectoral oversight, public bodies for the oversight of acquisition of AI systems for human rights sensitive use cases, such as criminal justice, law enforcement, welfare, employment, health care, among others, and independent judiciary systems.

64. Member States should ensure that governments and multilateral organizations play a leading role in guaranteeing the safety and security of AI systems. Specifically, Member States, international organizations and other relevant bodies should develop international standards that describe measurable, testable levels of safety and transparency, so that systems can be objectively assessed

and levels of compliance determined. Furthermore, Member States should continuously support strategic research on potential safety and security risks of AI technologies and should encourage research into transparency and explainability by putting additional funding into those areas for different domains and at different levels, such as technical and natural language.

65. Member States should implement policies to ensure that the actions of AI actors are consistent with international human rights law, standards and principles throughout the life cycle of AI systems, while demonstrating awareness and respect for the current cultural and social diversities including local customs and religious traditions.

66. Member States should put in place mechanisms to require AI actors to disclose and combat any kind of stereotyping in the outcomes of AI systems and data, whether by design or by negligence, and to ensure that training data sets for AI systems do not foster cultural, economic or social inequalities, prejudice, the spreading of non-reliable information or the dissemination of anti-democratic ideas. Particular attention should be given to regions where the data are scarce.

67. Member States should implement policies to promote and increase diversity in AI development teams and training datasets, and to ensure equal access to AI technologies and their benefits, particularly for marginalized groups, both from rural and urban zones.

68. Member States should develop, review and adapt, as appropriate, regulatory and legal frameworks to achieve accountability and responsibility for the content and outcomes of AI systems at the different phases of their life cycle. Member States should introduce liability frameworks or clarify the interpretation of existing frameworks to ensure the attribution of accountability for the outcomes and behaviour of AI systems. Furthermore, when developing regulatory frameworks, Member States should, in particular, take into account that ultimate responsibility and accountability must always lie with natural or legal persons and that 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 the actors and the technological processes involved across the different stages of the AI systems life cycle.

69. Member States should enhance the capacity of the judiciary to make decisions related to AI systems as per the rule of law and in line with international standards, including in the use of AI systems in their deliberations, while ensuring that the principle of human oversight is upheld.

70. In order to establish norms where these do not exist, or to adapt existing legal frameworks, Member States should involve all AI actors (including, but not limited to, researchers, representatives of civil society and law enforcement, insurers, investors, manufacturers, engineers, lawyers, and users). The norms can mature into best practices, laws and regulations. Member States are further encouraged to use mechanisms such as policy prototypes and regulatory sandboxes to accelerate the development of laws, regulations and policies in line with the rapid development of new technologies and ensure that laws and regulations can be tested in a safe environment before being officially adopted. Member States should support local governments in the development of local policies, regulations, and laws in line with national and international legal frameworks.

71. Member States should set clear requirements for AI system transparency and explainability so as to help ensure the trustworthiness of the full AI system life cycle. Such requirements should involve the design and implementation of impact mechanisms that take into consideration the nature of application domain (Is this a high-risk domain such as law enforcement, security, education, recruitment and health care?), intended use (What are the risks in terms of transgression of safety and human rights?), target audience (Who is requesting the information) and feasibility (Is the algorithm explainable or not and what are the trade-offs between accuracy and explainability?) of each particular AI system.

**POLICY AREA 3: DATA POLICY**

72. Member States should work to develop data governance strategies that ensure the continual evaluation of the quality of training data for AI systems including the adequacy of the data collection and selection processes, proper security and data protection measures, as well as feedback mechanisms to learn from mistakes and share best practices among all AI actors. Striking a balance between the collection of metadata and users' privacy should be an upfront goal for such a strategy.

73. Member States should put in place appropriate safeguards to recognize and protect individuals' fundamental right to privacy, including through the adoption or the enforcement of legislative frameworks that provide appropriate protection, compliant with international law. Member States should strongly encourage all AI actors, including private sector companies, to follow existing international standards and in particular to carry out privacy impact assessments, as part of ethical impact assessments, which take into account the wider socio-economic impact of the intended data processing and to apply privacy by design in their systems. Privacy should be respected, protected and promoted throughout the life cycle of AI systems.

74. Member States should ensure that individuals retain rights over their personal data and are protected by a framework which notably foresees transparency, appropriate safeguards for the processing of sensitive data, the highest level of data security, effective and meaningful accountability schemes and mechanisms, the full enjoyment of data subjects' rights, in particular the right to access and the right to erasure of their personal data in AI systems, an appropriate level of protection while data are being used for commercial purposes such as enabling micro-targeted advertising, transferred cross-border, and an effective independent oversight as part of a data governance mechanism which respects data sovereignty and balances this with the benefits of a free flow of information internationally, including access to data.

75. Member States should establish their data policies or equivalent frameworks, or reinforce existing ones, to ensure increased security for personal data and sensitive data, which if disclosed, may cause exceptional damage, injury or hardship to a person. Examples include data relating to offences, criminal proceedings and convictions, and related security measures; biometric and genetic data; personal data relating to ethnic or social origin, political opinions, trade union membership, religious and other beliefs, health and sexual life.

76. Member States should use AI systems to improve access to information and knowledge, including of their data holdings, and address gaps in access to the AI system life cycle. This can include support to researchers and developers to enhance freedom of expression and access to information, and increased proactive disclosure of official data and information. Member States should also promote open data, including through developing open repositories for publicly-funded or publicly-held data and source code.

77. Member States should ensure the overall quality and robustness of the dataset for AI, and exercise vigilance in overseeing their collection and use. This could, if possible and feasible, include investing in the creation of gold standard datasets, including open and trustworthy datasets, which are diverse, constructed on a valid legal basis, including consent of data subjects, when required by law. Standards for annotating datasets should be encouraged, so it can easily be determined how a dataset is gathered and what properties it has.

78. Member States, as also suggested in the report of the UNSG's High-level Panel on Digital Cooperation, with the support of the United Nations and UNESCO, should adopt a Digital Commons approach to data where appropriate, increase interoperability of tools and datasets and interfaces of systems hosting data, and encourage private sector companies to share the data they collect as appropriate for research or public benefits. They should also promote public and private efforts to create collaborative platforms to share quality data in trusted and secured data spaces.

## POLICY AREA 4: DEVELOPMENT AND INTERNATIONAL COOPERATION

79. Member States and transnational corporations should prioritize AI ethics by including discussions of AI-related ethical issues into relevant international, intergovernmental and multi-stakeholder fora.

80. Member States should ensure that the use of AI in areas of development such as health care, agriculture/food supply, education, media, culture, environment, water management, infrastructure management, economic planning and growth, and others, adheres to the values and principles set forth in this Recommendation.

81. Member States should work through international organizations to provide platforms for international cooperation on AI for development, including by contributing expertise, funding, data, domain knowledge, infrastructure, and facilitating collaboration between technical and business experts to tackle challenging development problems, especially for LMICs, in particular LDCs, LLDCs and SIDS.

82. Member States should work to promote international collaboration on AI research and innovation, including research and innovation centres and networks that promote greater participation and leadership of researchers from LMICs and other regions, including LDCs, LLDCs and SIDS.

83. Member States should promote AI ethics research by international organizations and research institutions, as well as transnational corporations, that can be a basis for the ethical use of AI systems by public and private entities, including research into the applicability of specific ethical frameworks in specific cultures and contexts, and the possibilities to match these frameworks to technologically feasible solutions.

84. Member States should encourage international cooperation and collaboration in the field of AI to bridge geo-technological lines. Technological exchanges/consultations should take place between Member States and their populations, between the public and private sectors, and between and among Member States in the Global North and Global South.

85. Member States should develop and implement an international legal framework to encourage international cooperation between States and other stakeholders paying special attention to the situation of LMICs, in particular LDCs, LLDCs and SIDS.

## POLICY AREA 5: ENVIRONMENT AND ECOSYSTEMS

86. Member States should assess the direct and indirect environmental impact throughout the AI system life cycle, including but not limited to, its carbon footprint, energy consumption, and the environmental impact of raw material extraction for supporting the manufacturing of AI technologies. They should ensure compliance of all AI actors with environmental law, policies, and practices.

87. Member States should introduce incentives, when needed and appropriate, to ensure the development and adoption of rights-based and ethical AI-powered solutions for disaster risk resilience; the monitoring, protection and regeneration of the environment and ecosystems; and the preservation of the planet. These AI systems should involve the participation of local and indigenous communities throughout their life cycle and should support circular economy type approaches and sustainable consumption and production patterns. Some examples include using AI systems, when needed and appropriate, to:

(a) Support the protection, monitoring, and management of natural resources.

(b) Support the prevention, control, and management of climate-related problems.

(c) Support a more efficient and sustainable food ecosystem.

(d) Support the acceleration of access to and mass adoption of sustainable energy.

(e) Enable and promote the mainstreaming of sustainable infrastructure, sustainable business models, and sustainable finance for sustainable development.

(f) Detect pollutants or predict levels of pollution and thus help relevant stakeholders identify, plan and put in place targeted interventions to prevent and reduce pollution and exposure.

88. When choosing AI methods, given the data-intensive or resource-intensive character of some of them and the respective impact on the environment, Member States should ensure that AI actors, in line with the principle of proportionality, favour data, energy and resource-efficient AI methods. Requirements should be developed to ensure that appropriate evidence is available showing that an AI application will have the intended effect, or that safeguards accompanying an AI application can support the justification.

**POLICY AREA 6: GENDER**

89. Member States should ensure that digital technologies and artificial intelligence fully contribute to achieve gender equality; and that the rights and fundamental freedoms of girls and women, including their safety and integrity are not violated at any stage of the AI system life cycle. Moreover, Ethical Impact Assessments should include a transversal gender perspective.

90. Member States should have dedicated funds from the public budgets linked to financing gender-related schemes, ensure that national digital policies include a gender action plan, and develop specific policies, e.g. on labour education, targeted at supporting girls and women to make sure girls and women are not left out of the digital economy powered by AI. Special investment in providing targeted programmes and gender-specific language, to increase the opportunities of participation of girls and women in science, technology, engineering, and mathematics (STEM), including information and communication technologies (ICT) disciplines, preparedness, employability, career development and professional growth of girls and women should be considered and implemented.

91. Member States should ensure that the potential of AI systems to improve gender equality is realized. They should guarantee that these technologies do not contribute to exacerbating the already wide gender gaps existing in several fields in the analogue world. This includes the gender wage gap; the representation in certain professions and activities gap; the lack of representation at the top management positions, boards of directors, or research teams in the AI field; the education gap; digital/AI access, adoption, usage and affordability gap; the unequal distribution of unpaid work and of the caring responsibilities in our societies.

92. Member States should ensure that gender stereotyping, and discriminatory biases are not translated into the AI systems. Efforts are necessary to avoid the compounding negative effect of technological divides in achieving gender equality and avoiding violence against girls and women, and all other types of gender identities.

93. Member States should encourage female entrepreneurship, participation and engagement in all stages of an AI system life cycle by offering and promoting economic, regulatory incentives, among other incentives and support schemes, as well as policies that aim at a balanced gender participation in AI research in academia, gender representation on digital/AI companies top management positions, board of directors, or research teams. Governments should ensure public funds (on innovation, research and technologies) are channelled to inclusive programmes and companies, with clear gender representation, and that private funds are encouraged through affirmative action principles. Moreover, policies on harassment-free environments should be developed and enforced together with the encouragement of the transfer of best practices on how to promote diversity throughout the AI system life cycle.

94. UNESCO can help form a repository of best practices for incentivizing the participation of women and under-represented groups on all stages of the AI life cycle.

**POLICY AREA 7: CULTURE**

95. Member States are encouraged to incorporate AI systems where appropriate in the preservation, enrichment, understanding, promotion and accessibility of tangible, documentary and intangible cultural heritage, including endangered languages as well as indigenous languages and knowledge, for example by introducing or updating educational programmes related to the application of AI systems in these areas where appropriate and ensuring a participatory approach, targeted at institutions and the public.

96. Member States are encouraged to examine and address the cultural impact of AI systems, especially Natural Language Processing applications such as automated translation and voice assistants on the nuances of human language and expression. Such assessments should provide input for the design and implementation of strategies that maximize the benefits from these systems by bridging cultural gaps and increasing human understanding, as well as negative implications such as the reduction of use, which could lead to the disappearance of endangered languages, local dialects, and tonal and cultural variations associated with human language and expression.

97. Member States should promote AI education and digital training for artists and creative professionals to assess the suitability of AI technologies for use in their profession as AI technologies are being used to create, produce, distribute and broadcast a variety of cultural goods and services, bearing in mind the importance of preserving cultural heritage, diversity and artistic freedom.

98. Member States should promote awareness and evaluation of AI tools among local cultural industries and small and medium enterprises working in the field of culture, to avoid the risk of concentration in the cultural market.

99. Member States should engage large technology companies and other stakeholders to promote a diverse supply and plural access to cultural expressions, and in particular to ensure that algorithmic recommendation enhances the visibility and discoverability of local content.

100. Member States should foster new research at the intersection between AI and intellectual property, for example to determine who are the rights-holders of the works created by means of AI technologies among the different stakeholders throughout the AI life cycle.

101. Member States should encourage museums, galleries, libraries and archives at the national level to develop and use AI systems to highlight their collections, strengthen their databases and grant access to them for their users.

**POLICY AREA 8: EDUCATION AND RESEARCH**

102. Member States should work with international organizations, private and non-governmental entities to provide adequate AI literacy education to the public in all countries in order to empower people and reduce the digital divide and digital access inequalities resulting from the wide adoption of AI systems.

103. Member States should promote the acquisition of “prerequisite skills” for AI education, such as basic literacy, numeracy, coding and digital skills, and media and information literacy, as well as critical thinking, teamwork, communication, socio-emotional, and AI ethics skills, especially in countries where there are notable gaps in the education of these skills.

104. Member States should promote general awareness programmes about AI developments, including on the opportunities and challenges brought about by AI technologies. These programmes should be accessible to non-technical as well as technical groups.

105. Member States should encourage research initiatives on the responsible use of AI technologies in teaching, teacher training and e-learning among other topics, in a way that enhances opportunities and mitigates the challenges and risks involved in this area. The initiatives should be accompanied by an adequate assessment of the quality of education and of impact on students and teachers of the use of AI technologies. Member States should also ensure that AI technologies empower students and teachers and enhance their experience, bearing in mind that emotional and social aspects and the value of traditional forms of education are vital in the teacher-student and student-student relationships, and should be considered when discussing the adoption of AI technologies in education.

106. Member States should promote the participation of girls and women, diverse ethnicities and cultures, and persons with disabilities, in AI education programmes at all levels, as well as the monitoring and sharing of best practices in this regard with other Member States.

107. Member States should develop, in accordance with their national education programmes and traditions, AI ethics curricula for all levels, and promote cross-collaboration between AI technical skills education and humanistic, ethical and social aspects of AI education. Online courses and digital resources of AI ethics education should be developed in local languages, especially in accessible formats for persons with disabilities.

108. Member States should promote AI ethics research either through investing in such research or by creating incentives for the public and private sectors to invest in this area.

109. Member States should ensure that AI researchers are trained in research ethics and require them to include ethical considerations in their designs, products and publications, especially in the analyses of the datasets they use, how they are annotated and the quality and the scope of the results.

110. Member States should encourage private sector companies to facilitate the access of scientific community to their data for research, especially in LMICs, in particular LDCs, LLDCs and SIDS. This access should not be at the expense of privacy.

111. Member States should promote gender diversity in AI research in academia and industry by offering incentives to girls and women to enter the field, putting in place mechanisms to fight gender stereotyping and harassment within the AI research community, and encouraging academic and private entities to share best practices on how to enhance gender diversity.

112. To ensure a critical evaluation of AI research and proper monitoring of potential misuses or adverse effects, Member States should ensure that any future developments with regards to AI technologies should be based on rigorous scientific research, and promote interdisciplinary AI research by including disciplines other than science, technology, engineering, and mathematics (STEM), such as cultural studies, education, ethics, international relations, law, linguistics, philosophy, and political science.

113. Recognizing that AI technologies present great opportunities to help advance scientific knowledge and practice, especially in traditionally model-driven disciplines, Member States should encourage scientific communities to be aware of the benefits, limits and risks of their use; this includes attempting to ensure that conclusions drawn from data-driven approaches are robust and sound. Furthermore, Member States should welcome and support the role of the scientific community in contributing to policy, and in cultivating awareness of the strengths and weaknesses of AI technologies.

**POLICY AREA 9: ECONOMY AND LABOUR**

114. Member States should assess and address the impact of AI systems on labour markets and its implications for education requirements, in all countries and with special emphasis on countries

where the economy is labour-intensive. This can include the introduction of a wider range of “core” and interdisciplinary skills at all education levels to provide current workers and new generations a fair chance of finding jobs in a rapidly changing market and to ensure their awareness of the ethical aspects of AI systems. Skills such as “learning how to learn”, communication, critical thinking, teamwork, empathy, and the ability to transfer one’s knowledge across domains, should be taught alongside specialist, technical skills, as well as low-skilled tasks such as labelling datasets. Being transparent about what skills are in demand and updating curricula around these are key.

115. Member States should support collaboration agreements among governments, academic institutions, industry, workers’ organizations and civil society to bridge the gap of skillset requirements to align training programmes and strategies with the implications of the future of work and the needs of industry. Project-based teaching and learning approaches for AI should be promoted, allowing for partnerships between private sector companies, universities and research centres.

116. Member States should work with private sector companies, civil society organizations and other stakeholders, including workers and unions to ensure a fair transition for at-risk employees. This includes putting in place upskilling and reskilling programmes, finding effective mechanisms of retaining employees during those transition periods, and exploring “safety net” programmes for those who cannot be retrained. Member States should develop and implement programmes to research and address the challenges identified that could include upskilling and reskilling, enhanced social protection, proactive industry policies and interventions, tax benefits, new taxations forms, among others. Tax regimes and other relevant regulations should be carefully examined and changed if needed to counteract the consequences of unemployment caused by AI-based automation.

117. Member States should encourage and support researchers to analyse the impact of AI systems on the local labour environment in order to anticipate future trends and challenges. These studies should investigate the impact of AI systems on economic, social and geographic sectors, as well as on human-robot interactions and human-human relationships, in order to advise on reskilling and redeployment best practices.

118. Member States should devise mechanisms to prevent the monopolization of AI systems throughout their life cycle and the resulting inequalities, whether these are data, research, technology, market or other monopolies. Member States should assess relevant markets, and regulate and intervene if such monopolies exist, taking into account that, due to a lack of infrastructure, human capacity and regulations, LMICs, in particular LDCs, LLDCs and SIDS are more exposed and vulnerable to exploitation by large technology companies.

**POLICY AREA 10: HEALTH AND SOCIAL WELL-BEING**

119. Member States should endeavour to employ effective AI systems for improving human health and protecting the right to life, while building and maintaining international solidarity to tackle global health risks and uncertainties, and ensure that their deployment of AI systems in health care be consistent with international law and international human rights law, standards and principles. Member States should ensure that actors involved in health care AI systems take into consideration the importance of a patient’s relationships with their family and with health care staff.

120. Member States should regulate the development and deployment of AI systems related to health in general and mental health in particular to ensure that they are safe, effective, efficient, scientifically and medically sound. Moreover, in the related area of digital health interventions, Member States are strongly encouraged to actively involve patients and their representatives in all relevant steps of the development of the system.

121. Member States should pay particular attention in regulating prediction, detection and treatment solutions for health care in AI applications by:

(a) ensuring oversight to minimize bias;

(b) ensuring that the professional, the patient, caregiver or service user is included as a “domain expert” in the team when developing the algorithms;

(c) paying due attention to privacy because of the potential need of being constantly monitored;

(d) ensuring that those whose data is being analysed are aware of and provide informed consent to the tracking and analysis of their data; and

(e) ensuring the human care and final decision of diagnosis and treatment are taken by humans while acknowledging that AI systems can also assist in their work.

122. Member States should establish research on the effects and regulation of potential harms to mental health related to AI systems, such as higher degrees of depression, anxiety, social isolation, developing addiction, trafficking and radicalization, misinformation, among others.

123. Member States should develop guidelines for human-robot interactions and their impact on human-human relationships, based on research and directed at the future development of robots, with special attention to the mental and physical health of human beings, especially regarding robots in health care and the care for older persons and persons with disabilities, and regarding educational robots, toy robots, chatbots, and companion robots for children and adults. Furthermore, assistance of AI technologies should be applied to increase the safety and ergonomic use of robots, including in a human-robot working environment.

124. Member States should ensure that human-robot interactions comply with the same values and principles that apply to any other AI systems, including human rights, the promotion of diversity in relationships, and the protection of vulnerable groups.

125. Member States should protect the right of users to easily identify whether they are interacting with a living being, or with an AI system imitating human or animal characteristics.

126. Member States should implement policies to raise awareness about the anthropomorphization of AI technologies, including in the language used to mention them, and assess the manifestations, ethical implications and possible limitations of such anthropomorphization in particular in the context of robot-human interaction and especially when children are involved.

127. Member States should encourage and promote 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. This should be done using multiple norms, principles, protocols, disciplinary approaches, and assessment of the modification of behaviours and habits, as well as careful evaluation of the downstream cultural and societal impacts.

128. Member States, as well as all stakeholders, should put in place mechanisms to meaningfully engage children and young people in conversations, debates, and decision-making with regards to the impact of AI systems on their lives and futures.

129. Member States should promote the accountable use of AI systems to counter hate speech in the online domain and disinformation and also to ensure that AI systems are not used to produce and spread such content, particularly in times of elections.

130. Member States should create enabling environments for media to have the rights and resources to effectively report on the benefits and harms of AI systems, and also to make use of AI systems in their reporting.

## V. MONITORING AND EVALUATION

131. Member States should, according to their specific conditions, governing structures and constitutional provisions, credibly and transparently monitor and evaluate policies, programmes and mechanisms related to ethics of AI using a combination of quantitative and qualitative approaches. In support to Member States, UNESCO can contribute by:

(a) developing a globally accepted methodology for Ethical Impact Assessment (EIA) of AI technologies, including guidance for its implementation in all stages of the AI system life cycle, based on rigorous scientific research;

(b) developing a readiness methodology to assist Member States in identifying their status at specific moments of their readiness trajectory along a continuum of dimensions;

(c) developing a globally accepted methodology to evaluate *ex ante* and *ex post* the effectiveness and efficiency of the policies for AI ethics and incentives against defined objectives;

(d) strengthening the research- and evidence-based analysis of and reporting on policies regarding AI ethics, including the publication of a comparative index; and

(e) collecting and disseminating progress, innovations, research reports, scientific publications, data and statistics regarding policies for AI ethics, to support sharing best practices and mutual learning, and to advance the implementation of this Recommendation.

132. Processes for monitoring and evaluation should ensure broad participation of relevant stakeholders, including, but not limited to, people of different age groups, girls and women, persons with disabilities, disadvantaged, marginalized and vulnerable populations, indigenous communities, as well as people from diverse socio-economic backgrounds. Social, cultural, and gender diversity must be ensured, with a view to improving learning processes and strengthening the connections between findings, decision-making, transparency and accountability for results.

133. In the interests of promoting best policies and practices related to ethics of AI, appropriate tools and indicators should be developed for assessing the effectiveness and efficiency thereof against agreed standards, priorities and targets, including specific targets for persons belonging to disadvantaged, marginalized and vulnerable groups, as well as the impact of AI systems at individual and societal levels. The monitoring and assessment of the impact of AI systems and related AI ethics policies and practices should be carried out continuously in a systematic way. This should be based on internationally agreed frameworks and involve evaluations of private and public institutions, providers and programmes, including self-evaluations, as well as tracer studies and the development of sets of indicators. Data collection and processing should be conducted in accordance with national legislation on data protection and data privacy.

134. The possible mechanisms for monitoring and evaluation may include an AI ethics observatory, or contributions to existing initiatives by addressing adherence to ethical principles across UNESCO's areas of competence, an experience-sharing mechanism for Member States to provide feedback on each other's initiatives, AI regulatory sandboxes, and an assessment guide for all AI actors to evaluate their adherence to policy recommendations mentioned in this document.

## VI. UTILIZATION AND EXPLOITATION OF THE PRESENT RECOMMENDATION

135. Member States and all other stakeholders as identified in this Recommendation must respect, promote and protect the ethical principles and standards regarding AI that are identified in this Recommendation, and should take all feasible steps to give effect to its policy recommendations.

136. Member States should strive to extend and complement their own action in respect of this Recommendation, by cooperating with all national and international governmental and non-governmental organizations, as well as transnational corporations and scientific organizations, whose activities fall within the scope and objectives of this Recommendation. The development of a globally accepted Ethical Impact Assessment methodology and the establishment of national commissions for the ethics of technology can be important instruments for this.

## VII. PROMOTION OF THE PRESENT RECOMMENDATION

137. UNESCO has the vocation to be the principal United Nations agency to promote and disseminate this Recommendation, and accordingly shall work in collaboration with other United Nations entities, including but not limited to the United Nations Secretary-General's High-level Panel on Digital Cooperation, the World Commission on the Ethics of Scientific Knowledge and Technology (COMEST), the International Bioethics Committee (IBC), the Intergovernmental Bioethics Committee (IGBC), the International Telecommunication Union (ITU), the International Labour Organization (ILO), the World Intellectual Property Organization (WIPO), the United Nations Children's Fund (UNICEF), UN Women, the United Nations Industrial Development Organization (UNIDO), the World Trade Organization (WTO), and other relevant United Nations entities concerned with the ethics of AI.

138. UNESCO shall also work in collaboration with other international and regional organizations, including but not limited to the African Union (AU), the Alianza del Pacífico, the Association of African Universities (AAU), the Association of Southeast Asian Nations (ASEAN), the Caribbean Community (CARICOM), the Caribbean Telecommunications Union, the Caribbean Public Services Association, the Common Market for Eastern and Southern Africa (COMESA), the Community of Latin American and Caribbean States (CELAC), the Council of Europe (CoE), the Economic Community of West African States (ECOWAS), the Eurasian Economic Union (EAEU), the European Union (EU), the International Association of Universities (IAU), the Organisation for Economic Co-operation and Development (OECD), the Organization for Security and Co-operation in Europe (OSCE), the South Asian Association for Regional Cooperation (SAARC), the Southern African Development Community (SADC), the Southern Common Market (MERCOSUR), as well as the Institute of Electrical and Electronic Engineers (IEEE), the International Organization for Standardization (ISO), and international financing institutions such as the World Bank, the InterAmerican Development Bank, and the African Development Bank.

139. Even though, within UNESCO, the mandate to promote and protect falls within the authority of governments and intergovernmental bodies, civil society will be an important actor to advocate for the public sector's interests and therefore UNESCO needs to ensure and promote its legitimacy.

## VIII. FINAL PROVISIONS

140. This Recommendation needs to be understood as a whole, and the foundational values and principles are to be understood as complementary and interrelated.

141. Nothing in this Recommendation may be interpreted as approval for any State, other social actor, group, or person to engage in any activity or perform any act contrary to human rights, fundamental freedoms, human dignity and concern for the environment and ecosystems.