



CEB
Chief Executives Board
for Coordination

High-Level Committee on Programmes (HLCP)

Input to the UNESCO Ad Hoc Expert Group
on the Ethics of Artificial Intelligence

(HLCP Virtual Consultation, 29 July 2020)

Input from the High-Level Committee on Programmes (HLCP) to the UNESCO Ad Hoc Expert Group on the Ethics of Artificial Intelligence

1. The first draft text of a recommendation on the ethics of Artificial Intelligence (AI) prepared by the Ad Hoc Expert Group (AHEG) is a solid foundation for the consultations on this important topic. The comments below are suggestions that can be considered by the AHEG during their work on the second draft in September 2020. These comments are of a general nature, while more specific comments have been submitted through the online consultation platform by individual UN entities.
2. We agree with the AHEG that the complexity of the ethical issues surrounding AI requires equally complex responses that necessitate the cooperation of multiple stakeholders across the various levels and sectors of the international, regional and national communities. This requires very clear and concrete aims and objectives outlined in the draft text of a recommendation. Therefore, we urge the AHEG to be more precise in Section II of the draft text, and to outline not only the overall aim, but also specific objectives of this normative instrument. The same is true for Section I of the draft text where the scope of application of the instrument seems to be too broad. Among others, without a clear and limited scope, it may be difficult to monitor implementation of the instrument. In addition, some of the recommendations in the draft text are addressed to different stakeholders and not only to Member States. Therefore, the document should be harmonized in this respect and it should be clearer where other stakeholders are addressed.
3. Several concepts in the draft text may need further clarification. Therefore, the AHEG might wish to consider unpacking them through either providing definitions or improving explanations thereof. In doing so, it must be kept in mind that some terms are not easy to define and might not have internally acceptable definitions, which is especially true for the notions of AI and ethics as such. Concepts such as cross-domain systems approach, fit-for-purpose governance responses and others require greater clarity as they are hard to translate cross-culturally and across languages. More work is needed to ensure that the terminology used is understood by all stakeholders, especially those who would be tasked with implementing the instrument, but also by marginalized and underrepresented groups. However, providing clarifications should be done without over-specificity to the current technological landscape as this document is only valuable if it remains relevant to unanticipated emerging technologies and developments in AI.
4. We support the approach of outlining values and principles in the draft text of a recommendation. Foundational values can serve not only as high-level guidance for stakeholders, but as necessary preconditions upon which closely interlinked principles and policy actions are to build. This recommendation will likely influence the interpretation of other relevant recommendations, guidance and instruments. In this regard, a clearer framing to distinguish values and principles would be helpful for implementers of the instrument. This would include a more actionable and definitive foundation from which to make recommendations about the ethical use of AI. We suggest that the AHEG considers a shorter list of values. When deciding on the set of values, it must be ensured that they do not overlap but rather add to each other and are situated on a relatively similar scale.

5. The current distinction between groups 1 and 2 of principles remains unclear and may need some further explanation or even rethinking the division as such. For example, it is not clear why the principle of fairness relates to group 2 (the property of the AI technology itself). The principles can benefit from more concreteness and targeted description. In particular, the principle of proportionality could be more clearly unpacked.
6. We commend that the principles and policy actions are firmly grounded in the international human rights framework. AI should be used to protect and promote the rights of people rather than infringe upon them, e.g. as the result of security management. It should be further stressed that new technologies need to provide new means to advocate, defend and exercise human rights, in line with the Secretary General's Call to Action. The recommendation would benefit from making the foundations of many of the proposed values, principles and policy actions in the international human rights framework more explicit. This would strengthen the legitimacy of these points, facilitate implementation and ensure coherence across the UN system. It would be useful to incorporate more explicit references to specific human rights norms and standards, such as civil, political, economic, social, cultural and labour rights throughout the recommendation. For example, when the recommendation addresses issues of discrimination, this could build on the language used in the relevant provisions of various human rights treaties. Moreover, the recommendation could encourage stakeholders to increase their engagement with a full range of international, regional and national human rights mechanisms.
7. We also welcome the identification of human dignity as the central value guiding the recommendation. Building on this, the recommendation could further unpack the inherent tensions emerging in an increasingly automated, AI-driven economic, political, social and cultural environment.
8. We also commend that environmental concerns are elevated to the level of values. Indeed, environmental concerns are often overlooked and significantly underrepresented in the AI ethics debate as compared to other ethical dimensions. This also includes such issues as environmental management of natural resources, preservation of biodiversity, combating climate change, ensuring animal welfare, and food safety, which are not necessarily covered by the notion of environmental well-being used in the draft text. This should not, however, disregard the potential of AI to contribute to sustainable development by providing technological solutions and preventing harm to the environment.
9. In the ethics of AI there is often an uncertainty as to how conflicts between ethical principles should be resolved (e.g. larger datasets to combat bias in AI versus privacy, avoiding harm versus accepting some degree of harm (risk-benefit), etc.). The AHEG should make it clearer what to do in the case there is a possible trade-off at stake. International human rights law, standards and principles provide invaluable guidance for addressing possible conflicts of interests and rights. The AHEG may wish to explore the role of social dialogue for these purposes, as it is a mechanism to identify criteria that can guide choices and can strike a balance between conflicting principles and policies. It must be noted that a prioritization or trade-off approach could run into risks and exacerbate vulnerabilities, especially of already vulnerable groups, such as migrants and children.
10. The United Nations Secretary-General's Roadmap for Digital Cooperation was launched after the draft text of a recommendation was prepared. We suggest that the AHEG explore whether their substantive

recommendations could be aligned with the four pillars laid out in the roadmap (calling for the use of AI that is “trustworthy, human rights-based, safe and sustainable, and promotes peace”). We appreciate that the drafters, in crafting the current “holistic framework of interdependent values, principles and actions”, were attempting to capture the complexity of ethics of AI and to avoid unhelpful over-simplification. However, the substance of the roadmap pillars may already be reflected in the draft text of a recommendation. In the interest of providing a clear and coherent vision to the international community, and in aligning UNESCO’s efforts with existing high-level guidance from the Secretary-General, it seems beneficial that the forthcoming standard-setting instrument would reflect the four pillars prescribed in the roadmap and would be aligned with the language thereof, if possible.

11. The preamble states that specific attention should be given to low-and middle-income countries, including but not limited to countries in Africa, Latin America and the Caribbean, and Central Asia, as well as Small Island Developing States. In addition, the AHEG may want to address more the diversity of country groups in terms of level and diffusion of AI technologies and control over them. The reason is that the relevance of particular ethical issues may not always be specific to income levels. For example, the concern for biased algorithms will be mainly an issue in specific sectors, which can be more relevant for high- and middle-income countries.
12. We second the endorsements of the need to actively include low-and middle-income countries in global AI governance, and the need to address diversity gaps and the digital divide are right to the point. However, the draft text of a recommendation could be more explicit and comprehensive in identifying the unique circumstances, challenges and opportunities that Global South countries face with regard to the proliferation of AI systems. For instance, this could be done by including a paragraph to that effect under one of the following sub-headings (or versions thereof): “Leaving no one behind” or “Diversity and inclusiveness”.
13. In addition, we believe there is a need for principles, actions and/or other recommendations aimed specifically at bridging gaps, capacities, resources, and power imbalances between Global North and Global South countries with regard to the proliferation of AI. For instance, recommendations relating to AI ethics in Global South countries might include that AI training data should be equitably sourced; that the benefits from the use of AI systems derived from and used in these countries should be equitably distributed; or that Member States in the Global South will require additional investment and assistance in education and capacity-building around AI. Further, policy actions 8 and 9 could further emphasize the need for Global South countries to lead (rather than merely being included in) international AI governance.
14. It is also important to leverage the development of AI systems in ways that ultimately benefit all countries and people. In this regard, the lack of legislation, e.g. on data privacy in some countries may create a legal vacuum, which could be used to implement AI operations that may be unethical. It should also be acknowledged that there are differences in the capacity to regulate between AI-producing and AI-consuming countries. Additionally, high-risk international contexts should be highlighted, particularly humanitarian contexts in which power asymmetries, technological experimentation, and donor interests can generate ethical issues related to new technology considering the extreme vulnerability of people in these situations. There is also a need to build guardrails to ensure that AI is consistent with human rights and is not used for criminal purposes, especially targeting the most vulnerable.

15. Greater emphasis is also needed to reflect the differential access to and impact of AI across subgroups and individuals, especially children, women, girls and youth, as well as migrants and displaced persons, which should run through the entire document. This also includes involvement in policies around AI. Youth should be meaningfully engaged as experts, creators, and decision-makers in shaping and implementing policies around AI. The document should also refer to “older persons” rather than “elderly” and “people” rather than “citizens”, which is much more inclusive. In addition, certain groups may require specific mention and coverage due to lack of full protection afforded by citizenship and who are particularly vulnerable as a result: migrants and displaced persons, incarcerated people, stateless people, indigenous people, etc.
16. The governance dimension can be further strengthened in the draft text of a recommendation as the efforts to develop AI ethics are intrinsically linked with it. This could include the role that the UN System could play to support the development of governance systems, whether at the national or multilateral level, as it is the governance system and a suitable legal framework that enables the implementation of ethics principles and determines the degree to which AI serves positive or negative objectives. International legal standards and norms should bind all relevant stakeholders including corporations. Furthermore, governance and rights-based regulation are essential for ensuring respect for human rights and addressing privacy concerns, especially in vulnerable communities to ensure that no one is left behind. The AHEG might also wish to explore the essential link between AI governance and data governance.
17. We acknowledge the inclusion of principles related to accountability. However, we urge that these principles more explicitly refer to the significance of ongoing monitoring and evaluation of AI throughout the entire lifecycle, rather than primarily focussing on the front end of the AI cycle. Critical reflection is required within the document and the exposition of the principles of the need for constant evaluation of the training and data sets, their relevance and the ongoing accuracy of AI algorithms with clear retirement or re-development of algorithms to ensure continued relevance and accuracy. But ongoing monitoring and evaluation of AI systems should also encompass aspects beyond a narrow focus on data sets and accuracy of algorithms. Negative and positive impacts of AI systems depend on the social, economic, political and legal context in which they are embedded; even accurate AI algorithms can have highly discriminatory effects if used in inappropriate contexts, for example when the deployment of AI-based automated decision-making in welfare contexts prevents people in need from having meaningful interactions with case workers.
18. Ethics of AI may also take into account the economic dimension and labour relations more. Thus, the economic dimension is currently not outlined within the aims of the draft text of a recommendation. Further, labour rights and economic, social and cultural rights more generally, can be emphasized separately within the human rights language throughout the text.
19. A generally stronger socioeconomic perspective, especially with regard to developing countries, can be outlined. Although “leaving no one behind” is included as a value, this perspective should be strengthened in the principles and policy action sections. The principle of solidarity can play a role in this regard. It can also have a larger significance in terms of international cooperation, notably supporting AI development and redistribution of the benefits of AI.

20. The principles sections on “diversity and inclusiveness” and “fairness” touch on some of the issues around inequalities and equity; however, they only mention demographic or social inequalities, while a broader spectrum and the interaction between different inequalities needs to be referenced therein. Among others, economic, race and gender inequalities embedded in or resulting from AI systems should be included in these principles or additional principles.
21. The policy action section should further address economic inequalities both between countries and within countries. These can include issues around equitable access to technology, infrastructure, data, etc. and the need to address certain tendencies towards monopolisation, given network effects. It can also include fundamentally unequal negotiating positions of different countries and communities.
22. We wish to note that expanding access to quality data is essential to ensure safe and trustworthy AI. Data governance and privacy policies need to be put in place that ultimately address ethical concerns. It is therefore important to emphasize that safe and trustworthy AI will be enabled by the development of sustainable, privacy-protective data access frameworks that foster better training of AI models utilizing quality data. One such example is the Global Data Access Framework.
23. In terms of legal frameworks, it would be important for the document to mention that existing frameworks can often be interpreted appropriately to provide guidance on artificial intelligence. What is crucial in that respect is the development of legislation on data protection and privacy as well as the existence of an independent data protection authority. Sometimes, appropriate laws and frameworks are already in existence and can be applied to AI and other new technologies as well. However, what is often lacking are a stronger accountability and transparency element, which should be highlighted in the text. In addition, the call to Member States to consider development and implementation of an international legal framework on this topic where this current normative instrument can be used as a basis. This should be in line with international human rights law.
24. The AHEG should strengthen the recognition that gender, racial and ethnic biases in AI and automation exist and are often intertwined, and if the technologies are not developed and applied in a gender-responsive way, they are therefore likely to reproduce and reinforce existing gender, racial and ethnic stereotypes and discriminatory social norms. If the data that the machine is presented with includes stereotypical concepts of gender, racial or ethnic bias, the resulting application of the technology will perpetuate such bias. The AHEG should also recommend that more women, other gender-diverse groups and gender experts, including from different ethnic origins, must be involved in the design and deployment of AI tools. AI should contribute to the fundamental value of leaving no one behind, especially vulnerable groups that include women and racial or ethnic minorities. It is important to highlight the importance of inclusion also of transgender and non-binary people. The AHEG should therefore further reinforce gender equality, diversity and inclusion to ensure that everyone benefits equally.
25. An aspect that has not been fully unpacked in the draft text is the ethics of AI vis-à-vis children. They are for the first time growing in an increasingly machine-mediated society with little evidence of the social development impacts this might have for them and their understanding of what society, human rights, freedom or autonomy means. Beyond the impact on existing human rights or the environment, AI will have an immense impact on how society perceives itself. Today’s children are at the forefront of this global AI experiment with little evidence of outcomes. In addition, children have no active voice

in the process and no reference point of what society was before AI. In reflecting on rights-based approaches explicit consideration should be given to the specific rights of the child, recognizing that they will have greater exposure to AI over the life course and their development mediated and filtered by AI through the information received, the opportunities afforded, the services received and by the overall impact of their unique physical and psychological attributes.

26. The current structure of policy actions is rather complex and prolonged to the extent that each Policy Action in fact consists of more than one policy action. This may require further work to simplify the structure and to reduce the number of proposed actions to the ones essential to achieving the aims and objectives of the normative instrument. Here we return to the comment above on making the aims and objectives of the document clearer and more concrete. This is important in order to keep the recommendation simple and pragmatic so that it can be operationalized quickly.
27. Impact assessments are critically important, and we welcome the attention given to it in the draft text. This notion is based on a broader concept of due diligence, which has been successfully used, e.g. to bring human rights accountability into the private sector. Therefore, AHEG might want to be more explicit about due diligence responsibilities of State actors and in the private sector in the draft text, or even consider including it as a principle as there is good practice demonstrating it can be a bridge for action. More generally, the dialogue with the private sector could be more explicit in the draft text. Furthermore, *ex ante* impact assessments can be used as important tools to prevent the misuse of technology. For example, in the field of food and agriculture there is need to understand the impacts of AI on smallholders and this could be done through targeted impact assessments that could also address socio-economic inequalities. The inclusion of ethical impact assessments specifically, proposed in the draft text, should be further elaborated and take a more prominent place in the document. The draft text would benefit from linking such assessments to human rights impact assessment methodologies.
28. Institutional capacity building aspects, present in the draft text, should be reinforced. This is particularly important for national, regional and international entities that are responsible for protection and adjudication in the case of violations, including standards and tools to be used.
29. Last but not least, we recommend the text to be ambitious in its aims. This includes making bold proposals and be more assertive in suggesting that a stronger international legal framework is needed. It is also important to be more explicit about the innovative features that are reflected in this draft text in order for it to stand out among other documents. Additional innovative ideas and proposals are highly welcome in the next version of the draft text.