

1.2.2019

## Feedback on the Draft ethics guidelines for trustworthy artificial intelligence

The objectives and goals set for the ethical and sustainable development and application of AI in society are well-contemplated, just and honourable. In the wider global view, it is also prudent to strive towards the branding of human-centric "Trustworthy AI made in Europe", as opposed to the mainly consumer-centric American AI or the government control focused Chinese AI. AI pursuits and developments are already embodied in a large number of everyday contexts. Now is the last moment to foster true reflection and discussion on an ethical framework for AI. This should be done also at the global level. Europe should be a driving force in discussing the ethics in AI internationally. Global rules, based on ethical values, should be established, and EU should be an initiator in this discussion. Principles and good practices adapted in the forthcoming final version of the "Ethic Guidelines for Trustworthy AI" will be a good starting point for the discussion on global ethic guidelines for AI. The integrating European Union has also previously set goals and expectations on the role of Europe at the global scale or level, the realisation of which has sometimes fallen short in implementation and been hampered by various nationalist policies. Thus, Europe has been in a disadvantaged position in the global competition. AI is developed everywhere with a knowledge-based set of technologies par excellence. It is thus imperative to coordinate joint European-level AI development efforts and not to be misled by assuming the other global players would not already have a technological edge and lead in AI. Notwithstanding, the ethical point of view of core principles and values is still precisely the correct and the best bet that Europe can bring to the table of discussion on AI at the global level.

AI, as most technologies as such, is value neutral. It can be used for either good or bad purposes. The usage and its ethicality depends on human beings developing and using AI. The fundamental rights described in the Chapter I are the basis for the Draft Document. This approach is highly favoured. Until now, the development of AI has been dominated by technological and commercial interests. Ethical concerns have been addressed once problems have arisen. Legislation and regulation always lag behind technical development. The Draft Document correctly points out we require "...guidance on what we should do with the technology for the common good rather what we (currently) can do with the technology". On the other hand, legislation must be renewed and updated so that it also enables the full potential of AI for good purposes. E.g. in the upcoming Copyright Package data mining should be enabled to wider extent than now is proposed. The ethical principles and correlating values presented in Section 4 are to be favoured. However, one should bear in mind that sometimes it might be difficult to determine, how to define e.g. "good" in a certain context. If someone's "good" is less "good" for someone else, whose good will be more respected? Values may also be in conflict with each

other sometimes. This is often the case e.g. when creating common security requires limitation on individual freedom. The AI HLEG asked for specific input on the Section 5. It should be noted that AI is already used for purposes that are in conflict with the fundamental rights or can be seen unethical. Identification technologies are already being used for identifying people and there are hidden attempts to affect on people's opinions and democratic elections with AI systems. Using AI for scoring citizens and societal control system is reality. Europe is not safeguarded from this kind of attempts, nor are the Europeans. A possibility of opting out is mentioned in the document, but it is not clear how this opt out would be made possible in large data sets collected by different public and private actors. However, the Europe should actively promote everyone's control over their own personal data, based on MyData approach. In terms of the longer-term concerns, it is highly likely that radical technological changes will take place in the future. For many concerns and risks presented in the document, the question is not whether but when will they realise. Legislation and conventions always lag behind technological development so it is wise to address all known concerns as they appear and be prepared also for unpleasant and unlikely scenarios. The history has shown that if something is possible, it will be used, unless it is regulated by international conventions. Examples of such successful conventions include e.g. prohibition of chemical and nuclear weapons, which have mainly been widely accepted and effective.

It is important to increase the education on data science in all levels of education. The expertise is yet not enough. In the modern world, all citizens must be given prompt training on digital skills as a part of the common knowledge. Everybody should also be taught to understand the value of their personal data. Data skills should be also included in the key skills of lifelong learning. It should be noted, that AI is based on code and algorithms written by people. Thus, data science studies should include ethics so that future experts will have understanding on ethic related issues in data handling and use of data. Non-discrimination is an important principle. It is important to understand that algorithms may cause unintentional harm in the real life, if potential risks or side effects are not recognised while creating algorithms. A case example of this might be an AI assisted system that ranks job applicants for an interview based on their applications. Unintentional discrimination may also remain unnoticed. Thus, it is important to carefully plan and simulate algorithms before taking them in use especially in public services.

Based on common values, the EU is a natural actor to promote ethical use of AI. The assessment list presented in the Draft is a good reference point when assessing Trustworthy AI. As the document states, the list is not exhaustive and assessment is a continuing process. At this stage, it remains unclear, if Trustworthy AI should be self-evaluated by an organisation, or if an external auditing should take place. In both cases, Trustworthy AI should have valuable and wanted status. There are many examples of national, regional, or international rating or certification systems that provide prestige to both organisations and consumers, such as Fair Trade, FSC



(Forest Stewardship Council) for sustainable forestry), and MSC (Marine Stewardship Council for sustainable fishing) certificates. Should there be a suchlike organisation to evaluate and credit organisations following the Trustworthy AI ethic guidelines? Such a system might promote ethical use of AI and encourage organisations, perhaps also other global actors, to make their systems more compatible with the Trustworthy AI principles. This has been experienced e.g. with the Bologna process in higher education, as non-EU countries have started reforms to adapt their systems to be more compatible with the European education system.

CSC – IT Center for Science is a Finnish center of expertise in information technology owned by the Finnish state and higher education institutions. CSC provides internationally high-quality ICT expert services for higher education institutions, research institutes, culture, public administration and enterprises to help them thrive and benefit society at large. CSC supports the EC's HLEG AI work to compile ethic guidelines for trustworthy AI and thanks for the opportunity to comment on the draft version and working document of the guidelines.