

High-level Panel Follow-up Roundtable 3C – Artificial Intelligence

Session 1: 16 December 2019, 10am-12pm EST

Meeting Note

Recommendation 3C: We believe that autonomous intelligent systems should be designed in ways that enable their decisions to be explained and humans to be accountable for their use. Audits and certification schemes should monitor compliance of AI systems with engineering and ethical standards, which should be developed using multi-stakeholder and multilateral approaches. Life and death decisions should not be delegated to machines. We call for enhanced digital cooperation with multiple stakeholders to think through the design and application of these standards and principles such as transparency and non-bias in autonomous intelligent systems in different social settings.

Champions

Government of Finland, Government of France, Future of Life Institute, UN Global Pulse, Office of the Special Advisor

Key Constituents

Arm, Article 19, Association for the Advancement of Artificial Intelligence (AAAI), Beijing Academy of Artificial Intelligence, Beijing Academy of Artificial Intelligence, Government of Canada, Centre for Artificial Research Intelligence (CAIR), Council of Europe, Counter-Terrorism Committee Executive Directorate (CTED), DATACTIVE, Element AI, Government of Ghana, Google DeepMind, Graduate Institute Geneva, IEEE, International Organization for Standardization (ISO), International Telecommunication Union (ITU), International Trade Law Division, OLA, Makerere University, Government of Malaysia, UN Office of Disarmament Affairs, OECD, OpenAI, Office of the President of the General Assembly, Oxford Internet Institute, Partnership on AI (PAI), Republic of Korea, Government of Slovenia, Government of Switzerland, The Future Society, Tsinghua University, UNESCO, UNICRI, Wadhvani Institute for AI, WIPO

Introductory Remarks from Champions

Recommendation 3C is uncontroversial and indisputable, but also very general. There is a lot already going on in this space so the question is what would be the value added of this group compared to what is already happening.

A very broad but key starting point is to maintain a human-centric approach that is built on trust. That means that AI systems will be at the service of human beings and not the other way around, but also that human rights are protected. We must have defined a framework, with personal data protection as a key consideration.

Attention also needs to be paid to coordinating our work with existing initiatives to create synergies and avoid duplication. There's also a lot of hard technical work that needs to be done to figure out how we

can make systems that are as robust, transparent, and accountable and to ensure that progress is made towards universality as fragmented legal regimes emerge.

Existing frameworks and discourse tend to focus on the risk of misuse of these technologies, but also important is the risk of not using these technologies for public good, not innovating fast enough and freely enough to be able to achieve the transformations needed to get to the 2030 agenda.

Themes raised by Key Constituents

Timing for AI collaboration: constituents repeatedly noted that the timing was right for more convergence between parallel initiatives which have emerged as the popularity and capability of AI has become more mainstream. Noting that there is an emerging political will, at both national and regional levels, for progress towards effective AI governance that enables innovation but prevents AI harms, through the use of both soft law and hard law.

Experience leading on AI principle development: It was noted that many participants had experience developing and achieving approval for AI principles in their institutions or organizations, including both governments and industry. This experience and expertise should be harnessed.

Process notes: many constituents talked about the type of Roundtable process they want to see unfold, with reference to transparency, inclusivity, capacity building, collaboration, human rights based, and non-duplicative, noting that over 200 ethical charters on AI have already been promulgated. It was the general opinion of the group that now is the time to “move from principles to practice” on AI

Joining technical and legal: participants wanted to see greater connection between technology development, technology standards and legislative work, with several AI research labs expressing their hope that Roundtable progress will be grounded in technical and research insights to help strengthen the implementation of the recommendation.

Auditing of AI: many constituents, including civil society, governments and private sector, expressed an interest in progressing on the issue of audits and approval of AI systems. Progress in this area will also be helpful for large-scale AI procurement by governments. Many different ideas on how to approach this topic were presented, many emphasizing explainability, with further discussion on who should be leading or responsible for these audits.

Capitalize on convening power of the Roundtable constituents: it was noted that the Roundtable group of about 30 voices, represented a much larger constituency. This is true for professional bodies that represent their constituents, private industry consortiums that represent many actors, or United Nations bodies that work in areas of frontier technology. This give the Roundtable group tremendous reach if there can be agreement on ways forward.

Participation in AI meetings: There are many AI summits, gatherings, conferences, and meetings that take place throughout the year and participants of this roundtable group should try to raise these

discussions through networks at these events to solicit feedback and engage with the widest possible constituency.

Measurement of AI progress: several constituents raise the concept that a regular measurement of AI progress, including the development of benchmarks, would be useful for policy-makers and developers. As AI capacity grows from extremely narrow to more general applicability, thorough experience with transparency, audit, and explainability will be essential.

Benefits to underserved communities: Application of AI tools to support countries and communities that would not normally reap the benefits of frontier technologies should be explored and pursued through capacity building alongside hardware and software support. This capacity building should be supported by technology companies who have not traditionally engaged with these communities at a global scale.

Peace and security: it was noted by several multilateral institutions that AI is closely linked to issues of international peace and security and that there are ongoing efforts through the United Nations to address this. The work of this Roundtable should be complementary to those processes.

Closing remarks

Champions acknowledged the calls for a transparent, rights-based and collaborative approach to implementing the recommendation and the need for universality rather than division or fragmentation. Noting again the diversity of the constituents on the call, with special appreciation for the technical AI capabilities represented.

Given the broad recommendation and the diverse set of constituents on the call, it was agreed that the Champions would propose to the Key Constituents the formation of a series of working groups to address specific challenges.

It was noted that AI will have an impact across many of the recommendations, including 1A, 1B, 3AB, and 4. It will be imperative that there is close collaboration among Roundtable groups and Champions to ensure complementarity.

Next steps

- Champions will circulate proposed Roundtable working groups for consideration and feedback from the Key Constituents