KRISTIN LITTLE (IEEE); Senior Public Affairs Manager, IEEE-SA
Re: "Public consultation – comments on the OECD Framework for Classifying AI"

The IEEE Standards Association (IEEE SA) acknowledges the OECD for its efforts to gather input on the preliminary findings of its Framework for the Classification of AI Systems. We are pleased to provide comments on the document.

As background, the IEEE SA, is a globally recognized standards-setting body within IEEE. We develop consensus standards through an open process that engages industry and brings together a broad stakeholder community. IEEE standards set specifications and best practices based on current scientific and technological knowledge. IEEE SA has a portfolio of over 1,500 active standards and over 650 standards under development, including technical and impact standards relating to Artificial Intelligent Systems, next generation networks, IoT and Cybersecurity.

Regarding the specific questions posed by the Committee, we offer the following responses:

1. Should core and non-core criteria be distinguished? I.e. should there be a core classification framework for information that is generally accessible and additional, more complex or technical, considerations?

Yes, the hierarchical structuring and presentation of the criteria will assist with comprehension and ordering. It also enhances the rationality of the framework that expedites its acceptance and implementation.

2. Which characteristics should be core criteria and which 'optional'?

One should avoid arbitrary classification. A methodology should be devised to arrive at a figure of merit/index for each criterion, both in terms of desirability and impact that treats risks and rewards equitably and fairly. One should then adopt a societal risk/reward approach in a given context to make the criteria Normative or Instructive/optional. The latter classification should be context-justified and not universal. In addition, IEEE SA strongly suggests that well-being indicators should not be optional, but rather, core criteria.

3. Can AI systems be classified consistently & reliably with the core criteria?

Doing so will provide a structured and consistent basis/start and all exceptions can be handled separately under a special class.

4. Which criteria should be in a more detailed, technically-oriented framework?

One should adopt the principle of proportionality and develop detail and decompose criteria based on their figure of merit for overall risk/reward. The precautionary principle should also be adopted for areas of extreme uncertainty.

5. Should there be industry or application domain specific criteria and classifications?"

Yes, the core criteria should be treated as a generic set that can be normalised/customised for specific context and domains of application. This would provide a consistent, rational, fair, and equitable framework addressing the generic and application/domain specific requirements as appropriate.

Some additional suggestions:

• In section I. 1) A. Industrial Sector, the paper references the International Standard Industrial Classification of All Economic Activities (ISIC REV 4). We would note that ISIC is an economic rather than an industrial classification of activities and services. However, it is a global, and relevant, framework for initial impact/relevance assessment.

• Regarding the discussion of optionality in paragraph 12, we would point out that it is also important to consider human In, On and Out of the loop with regard to the operation and outputs of the system.

• We suggest that Table 2, "Potential impact on select human rights and democratic values," should consider impact on ethical values as well as rights. The list of ethical values is more extensive and culturally sensitive than that of universal rights.

• On page 6, in "Figure 2 - Characteristics per classification dimension and key actor(s) involved" (see diagram below) "human rights and wellbeing" are listed on the bottom of the "context" list of items, with "industrial sector, business function, and critical function" listed first.

Figure 2. Characteristics per classification dimension and key actor(s) involved

• It is unclear from the diagram if the listed elements in each section are in a specific order, but we would strongly recommend that "human rights and wellbeing" be considered first versus last in the red "Context" section of the Figure. In the development of IEEE's publication, Ethically Aligned Design, First Edition (EAD), over 700 experts listed "Human Rights" as the first element to consider in AI design. Examining this first means that it will be considered prior to the manufacturing stage, not after (at which point it is harder to halt the process).

With this early consideration of human rights and wellbeing comes the opportunity for innovation and time savings, and the possibility for more sustainable and competitive prototypes. It will also help organizations better align with ESG and other legal requirements.

• We suggest defining the word beneficial at the outset. "Wellbeing indicators" can help with this. They use objective criteria / data about things like improvement in mental health for children or environmental restoration. Where "beneficial" is not defined, the default criteria for success is not standard except when using GDP or exponential growth criteria. This changes fundamental engineering and all other constraints. "Wellbeing" as IEEE EAD and IEEE standard 7010-2020 - IEEE Recommended Practice for Assessing the Impact of Autonomous and Intelligent Systems on Human Well-Being (and the OECD) define it, is not about "happiness" or "feelings" but provides the Key Performance Indicators for what is considered "success" with AI once deployed.

• Based on the above, for the fourth part of the cycle, it is strongly recommended that "end users" are added as "key actors." This is where "wellbeing" for things like mental health is critical and cannot be "optional." In the field of AI enabled toys / devices there is already plentiful evidence showing how children's mental health has been harmed. So where wellbeing is "optional," then an AI enabled toy may not physically harm a child, but could fundamentally harm their sense of agency, identity, and self worth.

• Similarly, noting that on page 15, in the section "H. Benefits and risks to wellbeing [optional criteria]" we strongly suggest that the items on this list, such as "mental health," "environmental quality" and "social connections" should not be optional criteria.

• Further, this section (section H) could benefit from consistent language for the principal concepts, generally "Benefits and Harms" are used together, and, for the uncertain events, "Risks and Rewards."

We commend the OECD for offering so much to the field of pragmatic wellbeing and wellbeing indicators for more than a decade. As OECD has the reputation of being one of the world's leading authorities in this area, we close by reiterating our suggestion that wellbeing be made a primary priority.