**Organisation for Economic Co-Operation and Development (OECD) Principles on Artificial Intelligence (AI)**

**(Adopted May 22, 2019 by OECD’s 36 member countries, along with Argentina, Brazil, Columbia, Costa Rica, Peru and Romania)**

**Section 1: Principles for responsible stewardship of trustworthy AI**

**RECOMMENDS**that Members and non-Members adhering to this Recommendation (hereafter the “Adherents”) promote and implement the following principles for responsible stewardship of trustworthy AI, which are relevant to all stakeholders.

**CALLS ON** all AI actors to promote and implement, according to their respective roles, the following Principles for responsible stewardship of trustworthy AI.

**UNDERLINES** that the following principles are complementary and should be considered as a whole.

**1.1.Inclusive growth, sustainable development and well-being**

Stakeholders should proactively engage in responsible stewardship of trustworthy AI in pursuit of beneficial outcomes for people and the planet, such as augmenting human capabilities and enhancing creativity, advancing inclusion of underrepresented populations, reducing economic, social, gender and other inequalities, and protecting natural environments, thus invigorating inclusive growth, sustainable development and well-being.

**1.2.Human-centred values and fairness**

a)AI actors should respect the rule of law, human rights and democratic values, throughout the AI system lifecycle. These include freedom, dignity and autonomy, privacy and data protection, non-discrimination and equality, diversity, fairness, social justice, and internationally recognised labour rights.

b)To this end, AI actors should implement mechanisms and safeguards, such as capacity for human determination, that are appropriate to the context and consistent with the state of art.

**1.3.Transparency and explainability**

AI Actors should commit to transparency and responsible disclosure regarding AI systems. To this end, they should provide meaningful information, appropriate to the context, and consistent with the state of art:

i.to foster a general understanding of AI systems,

ii.to make stakeholders aware of their interactions with AI systems, including in the workplace,

iii.to enable those affected by an AI system to understand the outcome, and,

iv.to enable those adversely affected by an AI system to challenge its outcome based on plain and easy-to-understand information on the factors, and the logic that served as the basis for the prediction, recommendation or decision.

**1.4.Robustness, security and safety**

a)AI systems should be robust, secure and safe throughout their entire lifecycle so that, in conditions of normal use, foreseeable use or misuse, or other adverse conditions, they function appropriately and do not pose unreasonable safety risk.

b)To this end, AI actors should ensure traceability, including in relation to datasets, processes and decisions made during the AI system lifecycle, to enable analysis of the AI system’s outcomes and responses to inquiry, appropriate to the context and consistent with the state of art.

c)AI actors should, based on their roles, the context, and their ability to act, apply a systematic risk management approach to each phase of the AI system lifecycle on a continuous basis to address risks related to AI systems, including privacy, digital security, safety and bias.

**1.5.Accountability**

AI actors should be accountable for the proper functioning of AI systems and for the respect of the above principles, based on their roles, the context, and consistent with the state of art.