World Academy of Science, Engineering and Technology
International Journal of Law and Political Sciences
Vol:17, No:06, 2023

Ethical Artificial Intelligence: An Exploratory Study of Guidelines

Authors : Ahmad Haidar

Abstract : The rapid adoption of Artificial Intelligence (AI) technology holds unforeseen risks like privacy violation, unemployment, and algorithmic bias, triggering research institutions, governments, and companies to develop principles of AI ethics. The extensive and diverse literature on AI lacks an analysis of the evolution of principles developed in recent years. There are two fundamental purposes of this paper. The first is to provide insights into how the principles of AI ethics have been changed recently, including concepts like risk management and public participation. In doing so, a NOISE (Needs, Opportunities, Improvements, Strengths, & Exceptions) analysis will be presented. Second, offering a framework for building Ethical AI linked to sustainability. This research adopts an explorative approach, more specifically, an inductive approach to address the theoretical gap. Consequently, this paper tracks the different efforts to have "trustworthy AI" and "ethical AI," concluding a list of 12 documents released from 2017 to 2022. The analysis of this list unifies the different approaches toward trustworthy AI in two steps. First, splitting the principles into two categories, technical and net benefit, and second, testing the frequency of each principle, providing the different technical principles that may be useful for stakeholders considering the lifecycle of AI, or what is known as sustainable AI. Sustainable AI is the third wave of AI ethics and a movement to drive change throughout the entire lifecycle of AI products (i.e., idea generation, training, re-tuning, implementation, and governance) in the direction of greater ecological integrity and social fairness. In this vein, results suggest transparency, privacy, fairness, safety, autonomy, and accountability as recommended technical principles to include in the lifecycle of AI. Another contribution is to capture the different basis that aid the process of AI for sustainability (e.g., towards sustainable development goals). The results indicate data governance, do no harm, human well-being, and risk management as crucial AI for sustainability principles. This study's last contribution clarifies how the principles evolved. To illustrate, in 2018, the Montreal declaration mentioned eight principles well-being, autonomy, privacy, solidarity, democratic participation, equity, and diversity. In 2021, notions emerged from the European Commission proposal, including public trust, public participation, scientific integrity, risk assessment, flexibility, benefit and cost, and interagency coordination. The study design will strengthen the validity of previous studies. Yet, we advance knowledge in trustworthy AI by considering recent documents, linking principles with sustainable AI and AI for sustainability, and shedding light on the evolution of guidelines over time.

Keywords : artificial intelligence, AI for sustainability, declarations, framework, regulations, risks, sustainable AI
Conference Title : ICAILEP 2023 : International Conference on Artificial Intelligence: Law, Ethics, and Policy
Conference Location : Toronto, Canada
Conference Dates : June 19-20, 2023