Abstract

Artificial intelligence has proved to be one of the most modern and revolutionary inventions in the history of humankind. With internet being widely available, everyone wants to advance towards automation. From giving less crowded directions on google maps to suggesting music on Spotify, AI indeed makes life much easier.

Recent studies have shown that AI has been the cause for many violations that includes privacy breach, biased results, excessive energy consumption, defying human rights and dignity. These intelligent algorithms are trained on historic data that might be skewed towards a particular gender, race or culture. Bodies like the EU have proposed their own principles on AI ethics. However researchers believe that such standards have to be designed in a way that it can be applied globally. UNESCO has come up with some recommendations for tracking AI ethics violations in the form of standards that a responsible AI should obey in all stages of its life.

Till now companies have been doing their own impact assessments which are rather partial towards their policies and agendas. However, it is equally important to monitor the actions of AI when it reaches the end users or is deployed in an environment.

This research aims at building an ontology named as AIIA-Ontology for the purpose of extracting and structuring the information in UNESCO document in order to use it as a model for conducting external impact assessment on a number of real world AI cases. This external assessment will help third parties that could also be end users to report an AI incident by examining its impact, underlying risks, and violations of set guidelines by UNESCO.

This report will explore the use of ontologies for representing any information in a domain of interest. Moreover it explains the entire development process of AIIA-Ontology along with few queries for information retrieval. It will also see a web application designed for conducting impact assessment.

By the end of this research it was learnt that ontology turned out to be an innovative and highly effective way for modelling UNESCO recommendations. The results of the impact assessment were also very satisfying in terms of AIIA-Ontology's ability to describe various AI incidents.