Extension of Moral Agency to Artificial Agents

Authors : Sofia Quaglia, Carmine Di Martino, Brendan Tierney

Abstract : Artificial Intelligence (A.I.) constitutes various aspects of modern life, from the Machine Learning algorithms predicting the stocks on Wall streets to the killing of belligerents and innocents alike on the battlefield. Moreover, the end goal is to create autonomous A.I.; this means that the presence of humans in the decision-making process will be absent. The question comes naturally: when an A.I. does something wrong when its behavior is harmful to the community and its actions go against the law, which is to be held responsible? This research's subject matter in A.I. and Robot Ethics focuses mainly on Robot Rights and its ultimate objective is to answer the questions: (i) What is the function of rights? (ii) Who is a right holder, what is personhood and the requirements needed to be a moral agent (therefore, accountable for responsibility)? (iii) Can an A.I. be a moral agent? (ontological requirements) and finally (iv) if it ought to be one (ethical implications). With the direction to answer this question, this research project was done via a collaboration between the School of Computer Science in the Technical University of Dublin that oversaw the technical aspects of this work, as well as the Department of Philosophy in the University of Milan, who supervised the philosophical framework and argumentation of the project. Firstly, it was found that all rights are positive and based on consensus; they change with time based on circumstances. Their function is to protect the social fabric and avoid dangerous situations. The same goes for the requirements considered necessary to be a moral agent: those are not absolute; in fact, they are constantly redesigned. Hence, the next logical step was to identify what requirements are regarded as fundamental in real-world judicial systems, comparing them to that of ones used in philosophy. Autonomy, free will, intentionality, consciousness and responsibility were identified as the requirements to be considered a moral agent. The work went on to build a symmetrical system between personhood and A.I. to enable the emergence of the ontological differences between the two. Each requirement is introduced, explained in the most relevant theories of contemporary philosophy, and observed in its manifestation in A.I. Finally, after completing the philosophical and technical analysis, conclusions were drawn. As underlined in the research questions, there are two issues regarding the assignment of moral agency to artificial agent: the first being that all the ontological requirements must be present and secondly being present or not, whether an A.I. ought to be considered as an artificial moral agent. From an ontological point of view, it is very hard to prove that an A.I. could be autonomous, free, intentional, conscious, and responsible. The philosophical accounts are often very theoretical and inconclusive, making it difficult to fully detect these requirements on an experimental level of demonstration. However, from an ethical point of view it makes sense to consider some A.I. as artificial moral agents, hence responsible for their own actions. When considering artificial agents as responsible, there can be applied already existing norms in our judicial system such as removing them from society, and re-educating them, in order to re-introduced them to society. This is in line with how the highest profile correctional facilities ought to work. Noticeably, this is a provisional conclusion and research must continue further. Nevertheless, the strength of the presented argument lies in its immediate applicability to real world scenarios. To refer to the aforementioned incidents, involving the murderer of innocents, when this thesis is applied it is possible to hold an A.I. accountable and responsible for its actions. This infers removing it from society by virtue of its un-usability, re-programming it and, only when properly functioning, re-introducing it successfully

Keywords : artificial agency, correctional system, ethics, natural agency, responsibility

Conference Title : ICOR 2021 : International Conference on Operations Research
Conference Location : Rome, Italy
Conference Dates : September 16-17, 2021