

Moral Reflection with AI – Necessary or Redundant?

Aishwarya Suresh Iyer
Technische Universiteit, Delft (TUD)

Abstract. Moral support with AI has been gaining traction. The proponents of moral support with AI claim that some of the more problematic behavioural patterns of humans can be resolved with the help of AI, such as inability to extend moral concern to global level problems like climate change and refugee crisis. They offer a variety of ways of doing so- through provision of more information, or helping them work through the procedural aspect of moral decision-making, or help them work through their normative positions. I disagree with the solution being offered because I don't see this as a problem which can be solved at an individual level. As the problems they want to fix are deep, systemic, institutional, socio-political problems, which may not be fixed by a moral support with AI system.

Keywords. Moral support with AI, climate change, biases

1. Introduction

Moral support with AI systems have been offered as theoretical possibilities to engage with human inability to be moral. Scholars such as Savulescu and Maslen [1], Klineciewicz [2], Frank [3], argue that our inability to be moral can be fixed with the help of AI systems.

In my poster I want to provoke discussions surrounding the creation of such a system. I will draw upon the ideas articulated here to design a poster. In the ensuing paragraphs, I will firstly expand upon the AI system proposed by Savulescu and Maslen [1], AI for moral enhancement. Secondly, I will expand upon the shortcomings in their proposal and argue that moral actions don't take place in a vacuum, they operate in a specific socio-political environment. So, our solutions may need to take into consideration the socio-political environment and not treat it as a moral problem alone. Finally, I conclude by reiterating the main points made in the previous paragraphs.

2. Heading

Savulescu and Maslen [1] argue that human psychology is limited by design. Our notions of morality and concepts related to it are limited by our psychology in that it has evolved in the context of smaller groups. People are influenced by immediate concerns such as people they perceive to be in their group, and are able to extend moral concern to them and fail to extend it effectively to people they perceive to be outside the group. So, we don't rationally deliberate, instead we are influenced by our biases.

Savulescu and Maslen [1] argue for moral enhancement with an AI system which can help us manage our biases better, by informing us of them and helping us overcome them.

Furthermore, they also maintain that the AI system can advise us on the right course of action based on our moral values in relation to issues like climate change.

There are a number of shortcomings with this approach. Firstly, it is not entirely clear why being influenced more by our immediate surroundings can give rise to ‘immoral’ action. For instance, driving a car to work instead of taking public transport or cycling to the place of work is seen as an action which is detrimental to the environment and contributing to climate change. If this person who drives to work, claims to care about the climate, then this person is going against their moral values and the AI system would try to locate the biases of this person and try to advise them on making a decision in consonance with their moral values. However, what if the person is making the decision to drive to work after taking in all the necessary information such as how much time it takes to drive to work, family commitments etc. The AI system proposed by Savulescu and Maslen [1] may say that under the influence of our immediate surroundings and problems, we have failed to take into consideration the harm we are producing by our action. I think that while they may not have factored in the harm they are producing by prioritising the family, I find it hard to call this an immoral action or even that they are biased.

Secondly, our moral responses don’t occur in a vacuum, they occur in a very specific socio-political-economic scenario. Take the same example of the person driving to work due to family commitments and time constraints. They feel that this is more pressing due to the structure in which they exist. If these are more affluent people who can afford to outsource caring for family, then it is easier for them to be consistent with their moral values and the AI system can help them in finding the right way of doing so. However, for a person who can’t access such resources, what can the AI system do? It may not be able to help them, as this is a structural problem and the solution lies in changing aspects of the structure, and that is not within the ambit of this AI system.

In conclusion, it is important to understand that our actions can’t be readily deemed as immoral for the context in which they operate is complex. Given that our actions can’t be readily deemed as immoral or biased, our reading of the moral limitations needs to be sufficiently deeper to understand the complexity of the problem. Our moral responses are not just informed by our limitations, biases and moral values we claim to hold, but also by many other pressing concerns and needs. Hence, we need to reflect more on what we are trying to fix. And also ask, can they really be fixed with gentle guidance and nudging from an AI system in the complex socio-political structure we exist in.

References

- [1] Savulescu, J. and H. Maslen (2015). “Moral Enhancement and Artificial Intelligence: Moral AI?” *Beyond Artificial Intelligence*, Springer: 79–95.
- [2] Klineciewicz, M. (2016) “Artificial Intelligence as a means to Moral Enhancement” *Studies In Logic, Grammar And Rhetoric* 48 (61) 2016 doi: 10.1515/slgr-2016-0061
- [3] Frank, L.E. (2020) What Do We Have to Lose? Offloading Through Moral Technologies: Moral Struggle and Progress. *Sci Eng Ethics* 26, 369–385 (2020).