

Can Artificial Ethical Assistants Enhance Human Decision-Making in Virtue Ethics?

Keywords: Artificial Intelligence, Virtue Ethics, Action – Guidance, Ethical Assistants

Abstract

In this paper I will present a conceptual framework that proposes the creation of Artificial Ethical Assistants. The proposed AEA will be grounded in a virtue theory of ethics. I will justify the creation of such an AEA by using existing literature on moral psychology and similar AI models. By critiquing existing proposed models ([Giubilini and Savulescu, 2018](#); [Lara and Rodríguez-López, 2024](#)), I will show how they either prove inadequate to effectively help us navigate the complexities of ethical life or they remain an unrealisable proposal with our current technology. This paper aims at proposing features that an AEA should have that are realisable in the foreseeable future. The proposed AEA will respect the user's autonomy and choices but remains sufficiently effective as to still warrant its use.

The theoretical groundwork for the AEA is largely based on Neo – Aristotelian virtue ethics (especially that of ([Annas, 2015](#))) as well as [Williams \(1985\)](#) conception of the ethical.

There are certain problems that Aristotelian theories of virtue face when it comes to moral decision making that an AEA could potentially help the phronimos resolve them. For example, the problem of moral relevance will be discussed. How AI systems can account for the problem of identifying morally relevant facts and how they could help us gather and consider morally relevant facts that we perhaps would not be able to do so otherwise.

In this paper I identify distinct problems that most versions of virtue ethics face, primarily regarding the identification and application of relevant virtues. I then argue how an AEA can

assist a person in navigating those problems with the hope of enhancing human cognitive and ethical limitations. I propose that the AEA should possess functions such as the ability to recommend relevant academic and literary readings as to potentially aid the users understanding of virtues in particular contexts; in accordance with Williams's conception of the ethical. Other functions of the AEA should include the collection and processing of data that is relevant to the user in a particular moral scenario when our own processes prove to be insufficient. It is however the case that current AI systems do not have the capability to solve certain ethical problems that a person faces when attempting to act virtuously in any sufficient degree and therefore, only the exercise of phronesis is warranted. The risks of how AI and similar technologies can have to our pursuit towards Aristotelian virtue are also given attention and this proposal attempts to show how worries of moral de-skilling ([Vallor, 2015](#)) might be accounted for by users and AEAs.

This paper concludes that AEAs can be used as to assist the phronimos in action – guidance without impeding his moral growth or restraining his autonomy in an unacceptable degree. At the same time the proposed AEA does require a high level of motivation and commitment from the user and will not act as a system that churns out moral recommendations that the user is expected to follow.

Annas, J. (2015). Applying Virtue to Ethics (Society of Applied Philosophy Annual Lecture 2014). *Journal of Applied Philosophy*, 32(1), pp. 1-14. <https://www.jstor.org/stable/26813442> [Accessed 2024/06/20/].

Giubilini, A. and Savulescu, J. (2018). The Artificial Moral Advisor. The “Ideal Observer” Meets Artificial Intelligence. *Philosophy & Technology*, 31(2), pp. 169-188. Available at: 10.1007/s13347-017-0285-z.

Lara, F. and Rodríguez-López, B. (2024). Socratic nudges, virtual moral assistants and the problem of autonomy. *AI & SOCIETY*. Available at: 10.1007/s00146-023-01846-3.

Vallor, S. (2015). Moral Deskillling and Upskilling in a New Machine Age: Reflections on the Ambiguous Future of Character. *Philosophy & Technology*, 28(1), pp. 107-124. Available at: 10.1007/s13347-014-0156-9.

Williams, B. (1985). *Ethics and the Limits of Philosophy*: Fontana.