Aristotelian Virtue Ethics as an Impasse-Resolver for the Discussion Surrounding Autonomous Vehicles

Karol Wapniarski[1]

[1]) Faculty of Psychology and Cognitive Science, Adam Mickiewicz University, Poznań, Poland
karol.wapniarski@amu.edu.pl

Applying Aristotelian Virtue Ethics to ethical discussions surrounding AI is a growing but still underrepresented trend. Several efforts have been made to examine the applicability of the Aristotelian system to AI-related issues, including defining the basic "AI virtues" to be implemented both at the level of AI research [4] and AI use [6], as well as questioning whether all Aristotelian requirements can, in fact, be satisfied by AI systems [3]. When it comes to implementing the "virtues" in practice, two approaches can be identified: either implementing them in AI systems themselves or designing AI systems in a way that helps humans cultivate them.

The purpose of my talk is to follow the first approach while applying the Aristotelian Virtue Ethics specifically to the debate surrounding Autonomous Vehicles (AVs) that has been active in the years 2016-2018, arguing that the Aristotelian system provides a theoretical framework that can help to resolve the otherwise unresolvable concerns around how the AVs should be programmed and behave given the possibility of road accidents.

In my previous work on ethical debates around AVs ([8]), I have argued that since the ethical discussion surrounding AVs can be conceptualized as an instantiation of the classical "Trolley Problem" dilemma, it cannot be satisfactorily resolved based solely on abstract ethical considerations (compare [7]). This, together with the fact that AVs impose a real and urgent problem to be solved (which thus needs to be solved somehow – see [5]), has led researchers to look for other means and to reduce the purely abstract discussion to more down-to-earth issues. The most pronounced trend in the subject literature has been the law-reductionism ([2], [9]), which answers the question of what AV should do in a case of unavoidable accident by referring to legislation and economic rationality; this, however, faces many internal difficulties, including the question of how new legislation concerning AV should be designed and by what standards the economic gains should be measured. Moreover, the research done by Bonnefon et al. ([1]) has shown that the discussion does not have a clear answer even when it is sought to be established through a moral contract, with the responses of people asked to choose between a pragmatic and deontological stance varying according to whether they themselves or their relatives are sitting in the car or not.

The result of this is that both purely pragmatic and purely deontological ethics cannot be satisfactorily applied. In the talk, I shall claim that the discussion can be resolved only when a third ethical stance is taken, namely that of an aretaic ethics or Virtue Ethics. I will argue that since an aretaic approach applied to AI systems themselves shifts the attention from abstract considerations to the actual performance of the system and its improvement, the situation

of a "Moral Dilemma" does not occur. Instead, the continual improvement of algorithms and the continual improvement as a principle of learning algorithms themselves (so as to reduce the risk of accidents occurring) becomes the most "ethical" stance.

References

[1] Bonnefon, J.-F., Shariff, A., & Rahwan, I. 2016. *The social dilemma of autonomous vehicles*. Science **352**, 1573–1576.

[2] Casey, B. 2017. *Amoral machines, or: How roboticists can learn to stop worrying and love the law*. Northwestern University Law Review **111**, 1347–1366.

[3] Constantinescu, M., Crisp, R. 2022. *Can Robotic AI Systems Be Virtuous and Why Does This Matter?*. International Journal of Social Robotics **14**: 1547–1557.

[4] Farina, M., Zhdanov, P., Karimov, A. et al. (2024) *AI and society: a virtue ethics approach*. AI & SOCIETY **39**: 1127–1140.

[5] Gogoll, J. & Müller, J.-F. 2016. *Autonomous cars: In favor of mandatory ethics setting*. Science and Engineering Ethics **23**, 681–700.

[6] Hagendorff, T. 2022. *A Virtue-Based Framework to Support Putting AI Ethics into Practice*. Philosophy & Theology **35** (55).

[7] Keeling, G. 2020. *Why trolley problems matter for the ethics of automated vehicles*. Science and Engineering Ethics **26**, 293–307.

[8] Wapniarski, K & Łupkowski, P. 2024. *Rational or emotional? Alan Turing's "Heads in the Sand" Objection and the discussion of Autonomous Vehicles*. BNAIC/BeNeLearn 2024, Pre-proceedings.

[9] Wendel, W.B. 2018. textitEconomic Rationality and Ethical Values in Design-Defect Analysis: The Trolley Problem and Autonomous Vehicles. California Western Law Review **55**, 129–163.