The ethical challenges facing engineers seeking to develop and deploy autonomous vehicles

Tornike Avaliani (ta503) – Gonville and Caius College

The 21st century has seen autonomous vehicles sold to the general public for the first time. Despite the limited autonomous capacity of these vehicles there is much to suggest that the autonomous capacity of future vehicles will expand considerably[1]. Experts anticipate numerous ethical challenges with developing and deploying autonomous vehicles, including but not limited to collision behaviour, privacy and cybersecurity challenges, liability disputes and inclusivity[2]. Autonomous vehicles raise stakes to a virtually un-paralleled level, as decisions made by computer programmes have the potential to sentence vehicle occupants or third parties to death. This poses a critical question: how do we ensure that programmes make ethical decisions? In this report I intend to outline the major ethical challenges facing engineers in seeking to develop and deploy autonomous vehicles and consider our capacity for managing these challenges.

Ethical discussions in the public sphere regarding autonomous vehicles have focussed in large part on an iteration of "The Trolley Dilemma", which is formulated approximately as:

"An autonomous car is barrelling down on five persons and cannot stop in time to save them. The only way to save them is to swerve and crash into an obstacle, but the passenger of the car would then die. What should the car do?" [3]

At face value this kind of dilemma appears to be the moral ultimatum that autonomous vehicles need to be capable of solving before they can be adopted into society. However, Bonnefon et al make a compelling case that this kind of thought experiment is good for nothing but raising public awareness of the ethical challenges involving autonomous vehicles[3]. This is because such a scenario is too distant from reality and misses a vital bigger picture issue. Instead Bonnefon suggests it is much more useful to consider autonomous vehicles as "implicit moral agents" rather than "explicit moral agents". That is to say that autonomous vehicles will not decide to sacrifice a party outright, but instead will decide who to put at greater risk[3]. The trolley dilemma should instead be replaced with "The Statistical Dilemma"[3]. In brief, the countless decisions automated vehicles will make to reduce the risk of collisions when aggregated over innumerable scenarios may lead to an unethical risk distribution across a large dataset (as opposed to an isolated incident such as in "The Trolley Dilemma"). Ultimately this is equivalent to "The Trolley Dilemma" but much more realistic in formulation.

Autonomous vehicles will no doubt feature AI or machine learning algorithms for decision making which have been shown to unintentionally incorporate bias and/or discrimination[4]. This means there is a very real ethical risk that autonomous vehicles could unintentionally and unethically put certain road users at an unreasonable or unfair risk. Subsequently, I expect those at greater risk to be

[1] General Motors 2021, GM Announces Ultra Cruise, Enabling True Hands-Free Driving Across 95 Percent of Driving Scenarios. Available at: https://investor.gm.com/news-releases/news-release-details/gm-announces-ultra-cruise-enabling-true-hands-free-driving [Accessed : 26/12/2021].

[2] Poszler. F, Geißlinger. M. (2021) "AI and Autonomous Driving: Key ethical considerations", Institute for Ethics in Artificial Intelligence

[3] IEEE, The Trolley, The Bull Bar, and Why Engineers Should Care About the Ethics of Autonomous Cars. Available at: https://ieeexplore.ieee.org/document/8662742 [Accessed : 31/12/2021]

[4] Springer Link, Measuring Discrimination in Algorithmic Decision Making. Available at: https://link.springer.com/article/10.1007/s10618-017-0506-1 [Accessed : 31/12/2021]

deterred from adopting autonomous vehicle technology, hindering the deployment of autonomous vehicles in the process. For example, if pedestrians are favoured too much in relation to autonomous vehicle occupants, consumers will simply choose to drive non-autonomous vehicles. Thus, reducing the potential market share for autonomous vehicles, deterring their development and deployment. Similarly, if occupants are favoured too much in relation to pedestrians, there could be an unacceptable and unethical risk to pedestrians' safety, leading to public rejection of autonomous vehicles and hence a deterring of development and deployment[5]. Therefore, it's vital that the engineers responsible for these algorithms and AI are conscious of the inherent risk of bias. Furthermore, there is an ethical risk that manufacturers may choose to develop algorithms that are deliberately biased in order to increase their potential market share. For example, if a manufacturer chose to distribute fatalities between pedestrians and occupants at one to nine in favour of occupants, they could increase their market share as consumers would feel safer in their autonomous vehicle model than in other models. This would most likely be considered vastly unethical and a significant barrier to the deployment and development of autonomous vehicle technology. Engineers can avoid such circumstances by working strictly within the ethical framework as outlined by the Royal Academy of Engineering and respecting the public good, which states holding paramount the health and safety of others as a key principle[6].

Societally we must decide what the acceptable distribution of risk and fatality across road users is and engineers must have a vocal role in these discussions. This will no doubt be a delicate process which will develop across decades and its details lie outside of the scope of this report. Nonetheless it will be fascinating to watch as public opinion shapes the way autonomous vehicles are developed and deployed.

In conclusion, there are an array of ethical issues which are associated with the deployment and development of autonomous vehicle technology. Arguably the most important of these is collision behaviour. On consideration of "The Statistical Dilemma", it becomes clear that it's imperative to the successful development and deployment of autonomous vehicles that the algorithms involved in collision behaviour are well aligned with public ethical standards and the distribution of risk across road users meets an acceptable ethical standard. This will likely involve government regulation to prevent unethical and discriminatory algorithms and AI as well as manufacturers deliberately developing discriminatory algorithms to increase their market share. Autonomous vehicles have rich potential to transform road safety and help in tackling climate change, without engineers making proper ethical considerations we may be unable to release this powerful technology.

---

[5] J. F. Bonnefon, A. Shariff and I. Rahwan. (2016) "The social dilemma of autonomous vehicles", *J. Sci.*, vol. 352, pp. 1573-1576.
[6] Royal Academy of Engineering, Statement of Ethical Principles 2017. Available at: https://www.engc.org.uk/media/2334/ethical-statement-2017.pdf [Accessed : 31/12/2021]