# Ethical Guidelines for Autonomous Vehicle

## Dhruval Bhatt

April 12th, 2020

Autonomous vehicles enable safer and more efficient travel by eliminating errors due to lapse in human perception and optimizing to road and driving conditions. As we prepare to relinquish some decision-making capacity to intelligent machines, prior to mass deployment of autonomous cars, many critical scenarios should be carefully considered, and appropriate guidelines need to be established. One such scenario is an event of a potentially fatal crash that could present the autonomous car with several choices and it must rely on pre-programmed machine ethics to act correctly. That is, in a manner acceptable to our society.

Currently, it is the manufacturer's responsibility to ensure quality of parts and production of a vehicle but once a car is sold, it is the customer's responsibility to ensure lawful and safe operation. In autonomous vehicles, liability should be reduced from the passengers to manufacturers, who need to ensure unbiased, accurate, responsive programming. The guidelines established by Germany's Ethics Commission for Automated and Connected Driving, treat autonomous vehicle performance as product liability and require public sector oversight through licensing and monitoring. Both are good steps for universal adoption. In addition, owners of the vehicles should be subjected to regulations that prohibit alterations that affect performance and/or overrides accepted guidelines. People who accept the ideology of philosophers such as Hobbes and Locke, who "held that all morally significant beings have certain rights, such as the right to life, liberty, and property", may resist the idea of reduced right to operation of their property (Moor 69). However, to balance the freedom to exercise their right and the necessity to fulfil their duty for the well-being of others, one must implicitly buy into social contract theory, that urges cooperation. Such as, accepting and following universally accepted guidelines for autonomous vehicles. If not, there should be a shift in liability and consequences from the manufacturer to driver.

Given that users of autonomous vehicles must give up their agency to decide how to react in a fatal car crash and pedestrians and other cars are subject to the decision made by this machine,

it is important to program decisions that reflect society's moral and ethical framework. It is imperative that the autonomous vehicles should be allowed only if the decision made in fatal event is as good as a human or better. Projects like Moral Machine form a good starting point to determine the baseline of what humans in different societies would do in different scenarios while thinking hypothetically. However, manufacturers as well as policymakers should seek large scale virtual reality based studies that subject diverse users to high pressure decision making in fatal crash situation to form a robust understanding of what humans would actually do and how does that fit in with theoretically driven ethics arguments.

In order to program potential preferences for possible scenarios to handle fatal car crashes, it is important to consider a few such scenarios. While driving, people may face situations that could either harm animal/property or a human. In that case, it would be appropriate to always save a human over an animal. From the results of Moral Machine experiment, it is evident that majority of humans would make that decision and it follows the notion of protecting morally significant beings that philosopher Bentham described as "any being that can experience pain and pleasure" (Moor 51). Some people would also want to protect animals from experiencing pain. However, in the framework of utilitarianism that looks at the net effect of an action and overall benefit of a rule, it is undeniable that the loss of a human life, that is intertwined with other lives through familial and economic bonds, would have a more negative impact. While deciding between animal or property, I propose saving the living creature, but it could be argued that the economic burden of the property loss could be high, and it would still be acceptable to choose the loss of animal life. This would need to be evaluated with society at large in consideration.

Such a decision becomes harder when choosing between one human and another. Should an autonomous vehicle be programed to save a human based on age, status in society, health condition or gender? While responses from Moral Machine Experiment indicate that many would agree to saving a child over an elderly or a woman over man, German ethics guideline forbids such discrimination. I, too, would agree with the treating all human lives equally. Kant's categorical imperative deems that one "should act on moral rules that … everyone else [follows] without deriving a logical contradiction" and treating people as ends (Moor 40). If in fatal crash situation, we start to decide who to save, we are agreeing that not all lives are equal and strip the respect for

some rational beings. In addition, even from a utilitarian calculation, with the limited perception of autonomous vehicle, it is difficult to evaluate the value of a person's worth and it is best to let it be a random selection. Expanding upon this idea that all humans are equal, it seems practical to save as many lives as possible. Therefore, when faced with a decision to save one life or many, it should be optimized to save more lives.

While autonomous vehicles will mitigate the individual moral agency and virtuous behavior, a programmed moral rule would help more people adhere to a conduct they would have liked to choose but may not in high pressure situations. Nonetheless, it is difficult to determine universally accepted ethical decisions. Different people and cultures may have differing emphasis, yet some frameworks can prevail with right reasoning. A global ethical framework for autonomous vehicles would be more difficult to achieve but not impossible if grounded in basic human values and built with global input. To do so, policymakers would need to abandon cultural relativism and ensure that it is an inclusive thought process and not just a Eurocentric reasoning. Then, respective societies can successfully champion it to appeal to local leaders and consumers.

Words: 999