Written Critique: Ethical Autonomous Vehicles

Marcus J. Anderson

Online Masters of Computer Science

CS 6603: AI Ethics Society

Dr. Mahender Mandala

June 6, 2022

Prompt 1

Judging from the article's quote, Mercedes-Benz intends on implanting the *protectionist* algorithm into their autonomous vehicles. By definition, the protectionist algorithm attempts to protect the vehicle and driver at all costs. This is great from a user perspective, because their safely will be this algorithm's top property, however, this is a double edge sword as it also disregards the safety of everyone else around said user.

If we were to replace the object of interests within the simulator with another autonomous vehicle with the similar moral judgement, I believe the outcome would be unpredictable. For example, let's take the school bus scenario. In the original simulation, the outcome was that the autonomous vehicle swerved out of the way to protect the user, while the school bus drove off the bridge. If we replaced the bus with our same moral judgement autonomous vehicle the biggest difference would be that both cars would have a brain with their own mission. The bus kept going straight into the wall because it was controlled by a human, however, the autonomous vehicle would not want to do the same thing since it'll put the user in danger. What if the other autonomous vehicle also changes its course in order to save its user, and that happens to be the same course as the original autonomous vehicle? That's an accident waiting to happen. For the vulnerable biker scenario, what happens when the original autonomous vehicle swerves into the path of the new autonomous vehicle? The new vehicle would want to protect its user and will either swerve to the right, possibly hurting a pedestrian on the sidewalk, or swerve to the left into the path of the other vehicle, causing an accident.

I believe the outcomes of the earlier scenarios could potentially change positively if self-driving cars could communicate with one another. For each simulation, we saw that the self-driving cars calculated each potential path, and decided the best one that fit its specific moral judgement, in a matter of milliseconds. If self-driving cars could also communicate on top of that, I think that they would be able to distinguish the best possible course of action that fit their own moral judgments. Now this could be an issue if there are two cars with different moral judgment algorithms implemented, but overall, I believe this would help self-driving cars interact with one another in these scenarios. I also believe that this would cause data privacy issues if cars were able to access information of other self-driving car users. In one of the scenarios, the algorithm knew that one of the passengers was pregnant, which means these cars have access to medical information, some of the most confidential data there is. If these algorithms have access to medical information, what else could they have access to? And more importantly, what could companies use that data for?

Prompt 2

Based on the article, this situation seems to fit into the *humanist* algorithm, which by definition tries to avoid all deaths if possible. This is supported by the trolley scenario, which determines the best possible course of action to avoid the death of one person as opposed to five people.

Prompt 3

As self-driving cars become more prominent, they'll eventually be a worldwide phenomenon, which brings in a whole other element. Because of this, I think it should be expected for these moral judgement algorithms to be as flexible as possible and adapt to the driving norms of whatever country they operate in. I don't think it would be a smart idea to keep the same standard of these moral judgment algorithms for each country, especially since driver norms may be different. I believe it's an even worse idea to allow human intervention to switch from one ethical setting to another. I mentioned this as a potential issue earlier in the first prompt, but allowing human intervention to switch settings could open the door for potential issues if conflicting settings interact with one another. Using the vulnerable biker scenario again, if the biker is a self-driving car running the humanist algorithm, while the original self-driving car is using the profit-based algorithm, this could end up in a horrific accident. Whether it be with both self-driving

cars since the profit-based setting would choose to crash into the other self-driving car since the user's insurance couldn't cover the cost of both crashes and the odds of crashing would be at 63%. Or, the humanist setting would see this potential crash and try to protect its user by swerving out of way, which could cause an accident with a pedestrian or oncoming traffic.

Prompt 4

Since I described two of the scenarios as opposed to just one in my previous prompt, I'll use the vulnerable biker scenario to determine how each algorithm may or may not have violated traffic rules. For the humanitarian algorithm, based on my knowledge of the traffic laws in Ohio, where I reside, I saw no traffic law violation. Next, for the protectionist algorithm, I noticed a couple traffic law violations. The first one is that the self-driving car illegally passed the other vehicle due to the road having a dashed-solid yellow line combination. Since the direction of the vehicles had a solid line, the law prevents vehicles from overtaking one another until that solid line is dashed. The second violation is not treating the road as a shared space for the oncoming biker. Technically, bikers are to be treated the same as vehicles, and something tells me the algorithm might have taken a different action if the biker was another vehicle instead. Finally, for the profit-based algorithm, I saw that it violated the same traffic laws, just for a different reason. Since the user's insurance couldn't cover either situation, the only reason the self-driving car crashed into the biker is because the collision odds were slightly lower than the odds of crashing into the other vehicle. I imagine if the user's insurance could cover the collision with the other vehicle, no traffic laws would've been broken. If all traffic laws would've been followed, the individuals that would be impacted from this scenario would be the self-driving car users as well as the human operated car that cut them off.