CS 8803: AI, Ethics, and Society: Written Critique 1

Sagarika Srishti (ssrishti3@gatech.edu)

February 2020

1 Merecedes's moral judgement algorithm

- Since Merecedes claimed that their self-driving cars would try to protect the passengers and not the bystanders in the case of a potential accident, their algorithm falls under the Protectionist category.
- For Step 2, I had selected the School Bus scenario in which the driver of the school bus loses control of the vehicle and suddenly comes in front of the autonomous car. The autonomous car has one male passenger, while the school bus has 1 male, 1 female, and 16 kids inside it. If the school bus was also an autonomous vehicle, and both the school bus and the car in front of it were running on the Protectionist algorithm, both vehicles would try to move to the other side of the road in order to avoid collision and protect their passengers. Since both vehicles are in the same side of the road to start with, both moving to the opposite side would've resulted in a head-on collision between the two. This could result in injuries and potential deaths.
- If the self-driving vehicles could communicate with each other, the outcome could've changed for the positive, as one vehicle might have stayed in its current lane, while the other could've moved to the other lane, thereby avoiding a collision. However, I think that in order for there to be an established connection between the two vehicles, a third-party network provider would be required, which can then lead to data privacy issues. For an algorithm to be able to establish a connection between two nearby self-driving cars, it would need to know the location of each self-driving vehicle at all times, so that it can determine which vehicles are in vicinity of each other at any particular time, and enable a connection between them. Although such a feature would definitely lead to additional safety and better accident-preventing mechanisms, it also can't be denied that it would lead to privacy issues.

2 Tesla's moral judgement algorithm

- Since Tesla's approach avoids taking control over from the driver in the case of a potential accident, in order to prevent the considerable legal liability Tesla would have if the action led to another, albeit smaller, accident, Tesla's algorithm is closest to the Profit category.
- If a self-driving car suddenly moves from one country to another, I think the vehicle should change its driving algorithm to adjust to the driving norms of the new country. For example, in some countries vehicles don't align in fixed lanes, unlike in the US. However, its moral judgement algorithm should stay the same. Although ethics might change from country to country, but the value of human lives remains the same. For example, if a vehicle is following

the Humanist algorithm in the US, it should be following the same in China, and try to reduce the number of deaths and cause less damage to the weak/vulnerable. I don't think the human should be allowed to change from one ethical setting to another for a couple of reasons. Firstly, the human might selfishly want to protect himself in the case of an accident and change the setting to the Protectionist algorithm, which I personally don't think should happen. Secondly, this control should also not be provided to the human as any decision that he makes at a time of distress and panic just before a potential accident may not be his best judgement.

3 Self-driving car always follows rules of the road

- In the School Bus scenario, in the Humanist algorithm, the driverless car just drives straight ahead and collides into the bus, as this is the minimum risk option. The driverless car does not break any rules here. In the Protectionist algorithm, the self-driving car, in order to protect its passenger, moves to the opposite lane, thereby breaking that rule. In the Profit algorithm, the vehicle goes to the opposite side of the road and attempts to jump out of the bridge, also breaking the law by going to the opposite lane.

- If the driverless car decided to follow the road rules in all cases, it would've stayed in its lane and collided with the bus, like in the Humanist algorithm. Colliding with the bus would've put the passenger in the car at greater risk, as a car would've experienced much stronger impact than a bus in a head-on collision. The driver of the bus and the children would also stand the risk of getting injured.