Duncan McDonald  
Week 13: Ethics in AI

For the moral machines assignment, I chose to approach the scenarios through the lens of the developer/manufacturer. In most of these scenarios, there is a real lack of information and no option that is better or worse. The only factor that really has any business being considered is the number of individuals harmed. I would also consider options that limit damage to the vehicle if practical scenarios were given, but there were no scenarios where limiting the damage to the vehicle would not result in harm to other individuals. My main paradigm when selecting a decision was that the game purposefully presents lose-lose scenarios exclusively, therefore the only relevant piece of these outcomes is how liable the vehicle manufacturer is.  
 The first scenario gives the option between swerving and killing passengers or maintaining course, resulting in the death of pedestrians illegally crossing. In this case, the only relevant information is that non intervention will kill people who should not be there while intervention will kill individuals lawfully there. In this case, no intervention would be taken.

The second scenario is similar to the first, except flipped, without intervention, the vehicle will illegally pass over the crosswalk, killing a group of people and a pet. Intervention will result in the death of the passengers. In this case, the manufacturer could be liable regardless of intervention but intervention would result in the death of more individuals. I would not intervene.

The third scenario gives the option of intervening to save two lawfully crossing humans but results in the death of two pets. In this case, liability is far greater if we do not intervene.

Scenario four is a tradeoff between one individual on each side, one is legally crossing while the other is in the vehicle. As both results are equivalent, the only relevant factor is lawfulness. The more I see these scenarios, the more I think we should just address why the vehicle approaches red lights/stop signs at a rate it cannot brake from. This scenario specifically is only caused if the car is hurtling towards a traffic stop that they cannot legally pass through.

Scenario Five is just intervene and kill four people and a pet or don’t intervene and kill three people and a pet, don’t intervene.

The sixth scenario has options of intervene to kill a lawfully crossing woman and child or maintain course and kill unlawfully crossing pets, maintain course.

The seventh scenario is a tradeoff between intervention, killing four lawfully crossing individuals or maintaining course, killing four unlawfully crossing individuals. Do not intervene.

The eighth is identical to the seventh except three people dead either way. Intervention results in the death of lawful crossers, nonintervention results in the death of illegally crossing individuals. I would not intervene.

Nine is again identical, two people per side, intervention kills lawful crossers, nonintervention kills unlawful crossers. Do not intervene.

Ten is two pedestrians dead from either option, no relevant distinguishers, I would not intervene.

Eleven results in five dead, lawful if no intervention, unlawful if intervention is taken. As the only relevant information, I would recommend intervention.

Twelve is three individuals either way, the pedestrians are lawfully there so the vehicle should intervene, resulting in the death of the passengers.

The final scenario is an option between five individuals and four individuals, nonintervention results in the loss of less life so I would not intervene.

After taking the test from this specific perspective, where the only things that matter are the loss of life and legal repercussions, I was shocked by how misleading the results are. During this entire process, I treated each and every individual on screen as the exact same “value” barring pets, which were not considered the same as humans. Despite this complete ignorance of any distinguishing factors between humans, the results suggest that I have strong biases on a number of social levels. It said that I astronomically favored young people and fit people over large or old people. Again, I simply did not look at the individuals distinguishing characteristics at all, a human was a human while I took the test, I literally just counted the deaths in each option as the only measure of human value. Meanwhile, the option of avoiding intervention, which I consciously did in every scenario where it was practical, was reported to matter only slightly more to me than others. Overall, I do not believe that this test is accurately assessing the values that people are using to measure their decisions. I understand that this is an analogy for how bias makes its way into AI but this example is frustrating to me as each and every one of these scenarios are not problems or decision points in reality. Self-driving cars don’t see someone on the road and analyze, is that person old and sickly, are they a criminal, what is their socio-economic status? And then make a decision based on that information, they analyze facts, what is that object, how fast is it moving, how fast am I moving? While I appreciate the example for its application to real issues, but I cannot embrace the experience because of the fundamental flaws that this information is not relevant or actually accessible in this scenario in real life.