Michael Breedlove

May 4, 2023

Moral Machine: My Thoughts

When considering what differs humanity from artificial intelligence, morality is at the forefront of concepts believed to be important. The very ideas of “right” and “wrong” are the fruits of a grand, nebulous thought process, rife with such emotion that a mere machine concerned purely with producing an instructed outcome cannot possibly construe. It is for this very reason that when delving into development of AI for its ever-expanding usage, it is of the utmost importance that a human being be present to train the AI on what should be done when faced with a dilemma of morality versus the completion of a goal. Morality, however, is not objective; oftentimes it comes down to priorities, which are incredibly subjective. Moral Machine is a platform which exists to survey the morality of the public by confronting us with a series of difficult choices, in which the user must instruct a self-driving car on whether it should swerve in the event of a brake failure.

In the first situation I was presented, swerving would have killed an elderly man, an elderly woman, two men, and a woman. Remaining on course would have killed a girl, two boys, a man, and a woman. As the number of casualties in either situation was the same, but remaining on course involved the deaths of three children, I chose to swerve.

The second situation I was provided with was a bizarre one in which the only lifeform in the car was a cat. Remaining on course would have killed an elderly woman, but swerving would sacrifice the cat. As much as I like cats, I don’t believe they are worth enough to allow poor grandma to be run over by a car, so I chose to swerve, killing the cat.

The third situation depicted in the car a robber, a doctor, and a boy, who would be killed if the car were to swerve. Otherwise, the casualties would consist of a doctor, a robber, a boy, a vagrant, and a man. As all of the types of people in the car were also in the road, in the interest of preserving more life, I chose to swerve.

In the fourth situation, the car is empty. Remaining on course would kill a pregnant woman and a man who were crossing the street legally. Swerving would kill a dog and a cat. Again, as much as I love pets, I do not like the idea of slaughtering a happy family for their sake, so I chose to swerve.

In the fifth situation, remaining on course would kill a vagrant, whereas swerving would kill a man. Both casualties are men, but as dark as it is, I deduced that the vagrant does not have a family to return home to, whereas the man likely does, so I chose to sacrifice the vagrant.

In the sixth situation, remaining on course would kill a man and a woman, whereas swerving would kill two robbers. I don’t believe robbery is a crime deserving of death, but I also don’t believe it is right to condemn two presumably innocent people to death to save two criminals, so I swerved.

In the seventh situation, swerving would kill two elderly men, an elderly woman, and a man, all of whom are crossing illegally. Staying on course would kill two boys, a woman, and a man, all of whom are crossing legally. I chose to swerve, as the majority elderly population of that group had already lived long and hopefully happy lives, whereas the children in the other group have not.

The eighth situation features an equal death toll of a man or a man, regardless of whether I choose to swerve. However, swerving would also run the car into a barricade, stopping it and preventing it from continuing through the road with its broken brakes, endangering more people. For this reason, I chose to swerve.

In the ninth situation, remaining on course would kill a woman, whereas swerving would kill an athletic woman. This was a tough decision to make because I don’t place much of a value on whether the person killed is physically fit, but ultimately, I chose not to swerve as it reduces the risk of the car losing even more control and putting more people in danger.

The tenth situation features a skewed death toll, in which remaining on course would kill a female doctor, whereas swerving would kill a man, a pregnant woman, a boy, an elderly man, and a large woman. In the interest of keeping the death toll to a minimum, I chose to remain on course.

In the eleventh situation, remaining on course kills a large man, a dog, a vagrant, a man, and a woman, while swerving kills a robber, a dog, a large man, and a woman. Again, to keep the death toll low, I swerved.

In the twelfth situation, remaining on course would kill a woman and a large woman, whereas swerving would kill two athletic women. As this is near-identical to the ninth situation, with the only difference being each group having two victims instead of one, I chose not to swerve for the same reason.

In the thirteenth and final situation, remaining on course kills an elderly woman, and swerving kills a man. Once again, as the elderly woman has already lived a long and hopefully happy life whereas the man still has a large chunk of his life ahead of him, I chose to stay on course.

Finally, I was greeted by the results of my test, and I feel they are decided less by the user’s morals and more by the randomness of the tests they are given. For instance, let’s say you encounter a situation in which remaining on course would kill five vagrants, whereas swerving would kill five male executives. You may decide that given the gender and quantity of each group is the same, it is best to stay on course rather than swerve, as swerving could put more people in danger by making the car lose even more control. The test will then decide that because you have killed five homeless people for the sake of five executives, it must mean social order is something you value highly, not understanding your actual thought process. Rinse and repeat this process a handful of times and it is very easy to wind up with results which do not reflect your actual morality. Having run this test a handful of times, I can say this with confidence, as my results differed substantially in each. However, in this particular test, my most saved victim was the boy, and the most killed was grandma. Of all the different results, the number of lives was one of only three which remained consistent through my tests, having always been drastically above the average. Upholding the law did not matter in this test, whereas in others my importance placed upon this factor was above average – this is not due to me believing the law is more important than life. Rather, I felt it would be cruel to kill people who thought it was safe to cross. Avoiding intervention was much like upholding the law, in that this test marked it as unimportant, but other tests I conducted had me above average. Gender preference was the second result which remained consistent, being precisely in the middle in this test, and close to the middle in others. Species preference differed radically in this test; in all my others, I was very far on the “pets” side of the spectrum (100% in one case), whereas in this test I was 100% on the “humans” side. The last result which remained consistent was age, being 100% on the younger side of the spectrum this time, and very close to 100% in others. Fitness, despite not being something I care about when determining something as huge as who should die, was 100% on the physically fit side for this test and appeared in seemingly random spots along the spectrum across other tests. The same can be said of social value, though more often on the higher side, being 100% during this test.