Leonardo Saveri
Annalisa Barbara

An Essay on Self Driving Cars (the moral machine[1])

THE MORAL MACHINE EXPERIMENT

The Moral Machine Experiment is a 'thought' experiment about what people think that a self-driving car should do in an emergency situation where the only two available options are to either save its passengers or to save the pedestrians that are crossing the street.
The fundamental idea behind the test is that the test-taker is not in the experiment. The test-takers need to choose as if they were the car. In no case, they were asked to choose between saving or sacrificing themselves or someone they know.
What makes this experiment interesting is the fact that it presents the test-taker with many different scenarios in which different subjects, number of people, social status, and laws are taken into consideration.

We will try to discuss the statistics about the results regarding the decisions that people would make in such a situation.
We have decided to focus our essay on this particular experiment seen in class since it gives us the possibility to explore different topics related to ethics combined with questions that are arising from the new disruptive technologies of these last years.
What criteria do you use to decide? What moral values and reasons?
In the end, we will try to give our personal "decision-making scheme" if we were to find ourselves in the situation of actually coding the self-driving car.

From the discussion made in class, it emerged that different people with different backgrounds had totally different opinions about what should the car prioritize in order to decide what to do.
For instance, for some people age was the most important factor to take into consideration because, from a utilitarian point of view, young people can give a more important and longer contribution to the world we live in. For other people, instead, laws were the first thing to analyze. If the pedestrians were obeying the law (i.e they were crossing the street with the green light) they had to be saved from the car. This decision was based on the idea that these people didn't deserve to die because they were completely innocent at the time of the situation. On the other hand, if pedestrians were crossing with a red light, the test-takers that decide to follow the law as the first criteria believe that the car should save the passengers.

But, is there a universal answer to the question: "who should be saved?" Is there a general criterion to follow?
All the people that were put in the situation of deciding who the car should choose to save, were affected by their subjective opinion that was influenced by social stereotypes that could influence their judgments, and for this reason, are people's feeling and reactions to these kinds of questions sufficient to justify ethical judgments? This is a common problem regarding ethical questions.
One of the most interesting results that we found in this experiment was the fact that people often choose to act rather than not, deciding not to let the car continue to move straight; this means that people tend to react actively when they are put in front of an ethical question.

We found out, discussing in class, that people having to choose to either save a kid or an elderly man/woman were not sure about what the car should do. About half of the subjects would prefer to save an elderly man while the rest of the group decided that a kid should be saved.
By looking at a much bigger sample than the class, the website shows us that the majority of people would save the younger subjects rather than the older ones following a utilitarian point of view. On the other hand, people from eastern countries tend to save the elderly probably because in these cultures older people are more respected.

We figured, based on the results that we obtained from this experiment, that people were biased by a social and *"right"* point of view. This bias was probably created by the fact that we are influenced by the society we are living

[1] https://www.moralmachine.net/

in and by the opinion of the people around us.
There is no correct answer to this question.
Another example where there was a lot of debate, was the one where the car was supposed to decide between saving a dog's life crossing the street, and a criminal stealing from an old woman and running away while crossing the street.
Based on the social perspective we could obviously say that the dog was innocent in this case and, also if he is not human, it didn't deserve to die in the car crash while the "bad guy", who was robbing the old lady, deserved it. On the other hand, he is still a human, and this makes the decision very conflictual because many people feel like a human life is more important than a dog's one.

The experiment was made by a survey based on other different situations and characters.
The subject had the opportunity to decide between either intervening, like moving or not moving in a dangerous situation, choosing related to the gender or fitness of the people on the street or in the car at the moment of the crash, considering humans and pets, the number of people and other factors.

In general, based on the statistical results, the majority of people tend to: save more lives rather than less, consider the laws rather than making a personal decision, save the women, save humans instead of pets, save younger instead of older, save more fit people rather than large and save people with higher status rather than lower (i.e. save doctors instead of thieves).

On average, the test-takers were indifferent about choosing to save the passengers rather than the pedestrians, showing us that usually other reasons are taken into consideration.
One of the most interesting things was that people usually chose to save more people rather than save just an individual.
So can this decision be made and still be considered unethical if all the members of a group agree on the decision? Probably not.
In fact, if we choose while in a group we can be biased by the judgment of others and something that in our mind is considered ethical, could be redefined in our mind because of the subjective opinion created by society itself.

The problem could be also considered not just from an ethical point of view but also from a philosophical perspective.
We based the philosophical stance on the point of view of one of the fathers of philosophy, Emmanuel Kant who once said that you should only act according to the "maxim whereby we can", and trust that this opinion will be defined and considered as a "universal law".
This statement is the opposite of the Utilitarianism principle, guided by the subjective opinions of others, and so that our judgment should be directed in a way that let us feel the less painful, because we can obviously say that in this situation no choice could bring us joy and happiness.

We can agree that technology made a big step forward in the future that we will live in, but there are some ethical questions that a machine just can't take, because while a code moves a machine, an ethical question is moved by a sentiment that only a human can feel in that situation.
We are convinced that only when a self-driving car will be able to feel emotions regarding choices made in dangerous situations, we could be free to go around trusting a machine to make decisions on its own that could potentially be fatal for some people.

Because being able to build a car that feels emotions is quite literally impossible, or at least very far from the present, we can only try to find a scheme that a possible engineer could implement in a car to help it make decisions. A possible way could be to organize a sort of decision tree based on the answer that a huge sample of people could give. Find some sort of pattern that could show how the majority of humans would answer such difficult, ethical questions and train the car to take the decision that most people would choose. This would require a huge amount of data and, while it would please most people, it would anger many others.

We want to conclude our essay by proposing some sort of scheme that, in our mind, a potential self-driving car could use.
The first thing that we would look for, is if either the passengers of the car or the pedestrians are pets. While many

would disagree, we do believe that in such a case the human should always be saved.
Secondly, if both of them are of the same species or have a mixture of them, we would look at the probabilities of the outcomes. It is almost impossible for two outcomes to have the same probability of having fatalities (just think about the fact that the passengers have an iron box around them). In this case, we would let the car choose the action that would have the least probability of causing casualties. If this is not possible, and, surely, either the passengers or the pedestrians will die, here are the following steps that we think the car should consider.
The car should follow the law, and save whoever is on the right side of it. If both are on the right side, it should then save the group that has the biggest number of people. If the number of people is the same, the car should save the youngest. This could be done by taking the mean of the ages of the people in the group. This could then become another critical question: "how should the car compute the age of a group of people?".
If the calculated age is the same, it should then save the group that has more women (another critical question: how do we define a woman? Is it who can procreate? Should we consider trans men as women?). Let's, for the sake of the argument, consider women people that can procreate. We would choose to save the group with most women because, from a utilitarian point of view, women can have kids, and those kids will be useful in the future.
Again, this is, in the end, another ethical question. Is it correct to save the women? What if the pedestrians are all the men on the planet and the passengers are all the women on the planet? They wouldn't be able to procreate, making it difficult to understand if, or who, to save.
Lastly, if all the aspects are the same, the car should choose randomly (another critical question, how could we reach total randomness?)

As we said in the beginning, it is impossible to give one right answer, but we tried to give a personal intake about the decision process of the car, diving into some of the more complex aspects of ethical decision-making.