RBE 500 Homework #7

Arjan Gupta

Discussion Essay

Please write a two-page essay (excluding references) discussing the existing and potential ethical issues in the application domain(s) of robotics you are interested in exploring in your graduate study and future career.

In this essay, I will be exploring the ethical concerns surrounding the advent of autonomous vehicles. Specifically, I will focus on self-driving vehicles in the civilian space, such as cars, buses, and trucks. As for a geographical and cultural focus, this essay will mostly take on a perspective encapsulated by transportation and traffic sensibilities within the United States of America.

The V. Müller report states that a problem qualifies as a problem for AI/robot ethics if we do *not* readily know what the right thing to do is, further causing us to not readily know the answers to certain questions. [4] There are several such unanswered questions with regard to level 5, full driving automation cars, as defined by the Society of Automotive Engineers. [2] Such vehicles will not even have steering wheels or acceleration/breaking pedals! [7] Let us ponder on that for a moment. I ask the reader to imagine themselves in a car that just moves, without the ability to override any controls. Personally, I am a big fan of cruise-control and steering assistance in my own car, but I am not sure I would feel comfortable sitting in a car without the ability to override its controls. While this feels like a small thought experiment, it is in fact a question of ethics. Acceptance of level 5 autonomous cars requires humans to 'trust' a machine, and just how averse will human society be towards this? Level 5 autonomy might become technologically possible, but will humans actually want to use it?

A survery conducted by Policygenius found that 76% of Americans feel less safe driving or riding in cars with self-driving features. [6] This seems like an astonishingly high number, but is perhaps understandable. Before any major technological advancement, humans have usually felt skeptical of the technology in question. Most humans likely felt unsafe traveling in commercial aircraft when it first came into existence. Yet, if we glance at all the ten debate topics outlined in the V. Müller report, only a few could explain this general insecurity around AVs (autonomous vehicles). After, besides a few corner-cases of morality (for example, the famous trolley problem), AVs are supposed to be designed to follow traffic rules, which intrinsically have very few 'grey areas'. One of these ten debate topics, however, appears as a general logical explanation — the opacity of AI systems. [4] Here, it is explained that the AI systems behind AVs are generally invisible to both the user and the programmer. It is therefore difficult to trust a 'black-box' system to take charge of moving heavy machinery that could endanger human life.

The antidote to this distrust seems to be multifaceted, however two major facets are widely available statistics on the safety of AVs, and technical education of the public. A well-cited,

comprehensive article from the AI Ethics journal lists among its many conclusions that AVs have the potential to increase traffic safety, and that people with high education levels are more positive towards AVs than people with lower education levels. [5] Therefore, as society begins to see more widespread studies and data on the safety of AVs, people might begin to trust AVs more. Furthermore, as education systems begin to focus more on STEM, the general understanding of AI will begin to increase.

An additional perspective is that the opacity of AI systems could eventually decrease its scope from being opaque to both programmer and user to just the user. For example, currently, most programmers view neural networks (NNs) as a black box. We understand that NNs have input, output, and middle layers, but the fact that the middle layers re-adjust their weights via backpropagation makes it extremely difficult to know why a fully-trained NN makes a prediction in a certain way. However a recent manuscript by Caglar Aytekin has shown that any neural network with any activation function can be represented as a decision tree. [1] Studies like this will give humans a deeper understanding of AI, and can evetually help users of autonomous vehicles battle against the 'fear of the unknown'.

Now, assume that the problem of the negative public perception of AVs is solved. There would still be many difficult ethical questions that either the manufacturer or the user would have to answer. Take for a moment one of the corner-cases of morality we mentioned earlier — the trolley problem. Any well-designed AV would have to account for the situation where, due to unpredictable environmental factors on the road, there would be rare situations where the AV would not be able to avoid a calamity. As an example, say there is suddenly a human in the path of the AV, but the vehicle does not have enough time to steer away while keeping both the passenger and pedestian safe. In this event, should the AV prefer the life of the passenger, or the pedestrian?

As humans, we always prefer saving our own life first. But an autonomous entity that prefers its 'owner' or passenger could quickly turn into a slippery slope. What would be the legal ramifications if the pedestrian was killed? Who would be responsible? Although the question was probably asked far differently, the Policygenius survey found that Americans are divided 50/50 on who should be held responsible if a car crashes while self-driving features are in control — the driver or the car manufacturer. [6]

The moral question of the trolley problem begins to enter the realm of artificial moral agents, as mentioned in the V. Müller article. [4] If AVs have to make life-and-death decisions at times, should they be regarded as having moral responsibilities, hence making them moral agents? For example, if the passenger is having a medical emergency, should a highly intelligent AV have the responsbility of detecting this and navigating to the nearest hospital? After all, less intelligent devices already offer emergency services, for example the crash detection feature by Apple. [3]. Furthermore, car manufacturers already have the means to listen to

our conversations and monitor our habits, as an Auto Week news article suggests. [8]. So, if a highly intelligent machine already has to make a life or death decision, why shouldn't it also monitor the wellness of its passengers? Let us look at this perspective a even deeper — suppose the passengers of an AV happen to be a kidnapper and their hostage. The existence of AMBER alert suggests that this does happen quite often. Suppose the AV can detect an abduction, should it be its responsibility to pull over and call emergency services? Of course this also detects with the privacy & surveillance debate topic mentioned in the V. Müller article. [4] And if AVs should not be allowed to store user's conversations in a remote location, then that forces the intelligent processing of the conversations to happen 'offline', or within the car itself. This would in fact reinforce the case for the AV being an artificial moral agent — because it can use its offline intelligence to ensure the wellness of its passengers.

AVs can expand into many other debate topics mentioned in the V. Müller article. [4]. To briefly mention another — machine ethics comes into question in the situation that an AV needs to break traffic rules knowingly. Perhaps, for example, the passenger has voluntarily indicated that they are having a life-threatening emergency and require to be transported to the hospital. It would be considered ethical for the AV to safely drive above the speed limit to get the human to help as soon as possible. This would mean that the machine would need to intelligently decide when to break the law.

References

[1] Caglar Aytekin. Neural networks are decision trees, 2022.

[2] On-Road Automated Driving (ORAD) Committee. *Taxonomy and Definitions for Terms Related to Driving Automation Systems for On-Road Motor Vehicles*, April 2021.

[3] Tech Crunch. Apple offers a deeper dive into crash detection. https://techcrunch.com/2022/10/10/apple-offers-a-deeper-dive-into-crash-detection/, October 2022. Accessed: 2022-11-18.

[4] Vincent C. Müller. Ethics of Artificial Intelligence and Robotics. In Edward N. Zalta, editor, *The Stanford Encyclopedia of Philosophy*. Metaphysics Research Lab, Stanford University, Summer 2021 edition, 2021.

[5] Kareem Othman. Public acceptance and perception of autonomous vehicles: a comprehensive review. *AI Ethics*, 1(3):355–387, February 2021.

[6] Policygenius. Self-driving cars make 76% of americans feel less safe on the road. https://www.policygenius.com/auto-insurance/self-driving-cars-survey-2022/, September 2022. Accessed: 2022-11-18.

[7] Synopsys. Autonomous driving levels. https://www.synopsys.com/automotive/autonomous-driving-levels.html. Accessed: 2022-11-18.

[8] Auto Week. Your car is not necessarily listening to you, but it's definitely paying attention. https://www.autoweek.com/news/technology/a1708441/your-cars-radio-may-be-listening-you/, 2018. Accessed: 2022-11-18.