Brian Sitenga
15 November 2020

Exploring Ethical Decision Making of Artificial Intelligence in Vehicles

Abstract

The ethics of artificial intelligence is a growing topic of concern given the astronomical rise of technology that uses it. Given the expected rise of market demand for autonomous vehicles, or AVs, (Autonomous Vehicle Market Size 3), the ethical implications of truly autonomous driving came under greater scrutiny in the past decade. Enter Moral Machine (Rahwan 1), a platform that gathered public opinion of moral decisions made by AVs. It saw extreme success in the media and its creators wrote several scientific papers studying ethical decision making patterns across different demographics. However, the creators of Moral Machine made several errors in their methodology, and their results should serve only as an interesting thought experiment and not as substantial research. I created a new platform, Ethics Engine, that addresses these errors and collected data for the purpose of studying ethical decision making. This paper will refute Moral Machine's results as socially significant and discuss the data gathered from Ethics Engine.

Currently, Moral Machine presents users with randomly generated moral dilemmas in which an AV must either hit pedestrians in front of it or swerve into pedestrians in the adjacent lane. Essentially, it presents users with different variants of the Trolley Problem (Thomson 4). However, Moral Machine focuses less on utilitarianism and more on the social value or actions of the pedestrians in its way. For example, a group of pedestrians could include doctors, children, animals, criminals, pregnant women, larger people, and/or the elderly. Instead of focusing on the ethical implications of autonomous driving, the questions shift the focus onto how we value different social demographics. This study of social bias, while interesting, is problematic.

Moral Machine's results show that the general public prefers to save younger children, men, and the fit. These results are useless for making recommendations for public policy. Imagine enacting into law that programmers and vehicle manufacturers must create cars that swerve exclusively into elderly women. Clearly, social bias has no place in AI, and doing so would target and punish specific demographic groups. Additionally, the questions themselves have no practical value. AI in vehicles is not trained to recognize social status, age, or previous criminal history. It simply recognizes whether or not there is a pedestrian in the way. Therefore, the questions that Moral Machine poses are not meant to study the ethics of autonomous vehicles. Rather, they are hypothetical thought experiments about which groups of people we value more in society. Thus, Moral Machine's flawed results are a symptom of asking the wrong questions and collecting biased data.

Ethics Engine was created to address these problems and study more practical ethical issues surrounding AVs. First, no distinctions are made between any of the pedestrians in each question, removing social bias from the equation. Second, the questions expand their focus beyond just utilitarianism to study the different ethical ramifications of AVs swerving into a different lane. The primary issues being studied are utilitarianism, action vs. inaction, passengers vs. pedestrians, and known vs. unknown. The questions were methodically generated, and each user was asked the same set of questions, ensuring that all values were studied equally among participants. This paper's discussion of results may be updated as the number of survey responses grows and as the average preferences and values change over time. The current data collected from 73 participants will be used as a representative group of people living in North America.

Results and Discussion

The following preferences are ranked from a scale of 0 (least chosen) to 1 (most chosen) from the average of all participant responses:

Saving More - .49 vs. Saving Less - .40

Action - .48 vs. Inaction - .51

Known - .41 vs. Unknown - .59

Pedestrians - .59 vs. Passengers - .43

From this, we can see that participants had a fairly heavy preference for choosing options that saved more lives in total. Across all questions, they were 22.5% more likely to choose the option that resulted in fewer deaths. Ignoring all other factors, taking the utilitarian route is simple and effective. You simply tell the car to take the action that saves the most lives, not taking into account whose lives are lost and in what way. Utilitarianism has its downsides, but it could be a useful metric for determining public policy.

There was very little preference for taking action or being inactive. Participants were 6.3% more likely to stay the course, but this is not statistically significant enough to mark it as a preference. Survey takers were much more likely to choose an answer based on one of the other factors. For example, when given a choice between a known consequence and an unknown consequence, users very heavily preferred unknown consequences. People were more likely to take the risk with unknown choices 43.9% of the time. I was surprised that most participants felt this strongly, especially considering that the unknown option was often to swerve into the opposite direction of traffic. I left these questions open to interpretation, so many participants might have felt that it was worth taking the chance that there were no cars in the opposite lane, saving everyone. I also believe that it comes from a place of compassion. Very few people

wanted to cut their losses and take a life, especially if it might not have been necessary. Finally, there's another heavy preference, this time for pedestrians. People were more likely to save pedestrians over passengers 37.2% of the time. This provides useful insight into how we should design autonomous vehicles. Passengers should be willing to take the risk and responsibility of being in a car and should prioritize protecting pedestrians.

These issues weren't split as evenly as I predicted. In fact, most users felt very strongly about particular issues and answered the survey questions accordingly. Because these issues were so decisive, it becomes even more important to ensure that these concerns are heard by policymakers and companies that design autonomous vehicles.

References

"Autonomous Vehicle Market Size, Share: Industry Report, 2030." Autonomous Vehicle Market Size, Share | Industry Report, 2030, www.grandviewresearch.com/industry-analysis/autonomous-vehicles-market.

Crockett, Molly. "The Trolley Problem: Would You Kill One Person to Save Many Others?" The Guardian, Guardian News and Media, 12 Dec. 2016, www.theguardian.com/science/head-quarters/2016/dec/12/the-trolley-problem-would-you-kill-one-person-to-save-many-others.

Goh, Brenda, and Yilei Sun. "Tesla 'Very Close' to Level 5 Autonomous Driving Technology, Musk Says." *Reuters*, Thomson Reuters, 9 July 2020, www.reuters.com/article/us-tesla-autonomous/tesla-very-close-to-level-5-autonomous-driving-technology-musk-says-idUSKBN24A0HE.

Goldin, Pete. "10 Advantages of Autonomous Vehicles." *ITSdigest*, 2018, www.itsdigest.com/10-advantages-autonomous-vehicles.

Jaques, Abby. "Why the Moral Machine Is a Monster." *Miami.edu*, robots.law.miami.edu/2019/wp-content/uploads/2019/03/MoralMachineMonster.pdf.

Maddox, Teena. "How Autonomous Vehicles Could Save over 350K Lives in the US and Millions Worldwide." *ZDNet*, ZDNet, 1 Feb. 2018, www.zdnet.com/article/how-autonomous-vehicles-could-save-over-350k-lives-in-the-us-and-millions-worldwide/.

Maxmen, Amy. “Self-Driving Car Dilemmas Reveal That Moral Choices Are Not Universal.” *Nature News*, Nature Publishing Group, 24 Oct. 2018, www.nature.com/articles/d41586-018-07135-0.

Rahwan, Iyad, et al. *Moral Machine*, 2015, www.moralmachine.net/.

Thomson, Judith. “The Trolley Problem.” *Vanderbilt.edi*, 1984, www.psy.vanderbilt.edu/courses/hon182/thomsontrolley.pdf.