Sahib Bajwa

Mullen

CSCI 3202

13 November 2020

Short Paper on Ethics and AI

The autonomous vehicle industry is growing faster now than it ever has before. The history of the autonomous vehicle allows us to see the rapid growth of the industry. It also allows us to envision where the industry will be in the near future. The autonomous vehicle industry has grown for the manufacturers, shareholders, and consumers. There are now more options than ever for customers to choose from when selecting a new autonomous vehicle. Moreover, autonomous vehicles are becoming more appealing to customers due to their promising future at reducing crash statistics. This is due to the potential that autonomous vehicles have in the future, and have recently started showing, at being safer on roads by removing the element of human error from driving. Human error is the leading cause for vehicle crashes and injuries in said crashes, and with autonomous vehicles having a chance at helping people stay safer, they have become more appealing to customers. Autonomous vehicles contain powerful technology in both their hardware and software. These technologies help the car function with high certainty of success and helped them become safer and smarter over time. With all this innovation, autonomous vehicles have become more appealing to customers looking at cars with new features and better safety ratings. With this growth however has come a lagging behind by all autonomous vehicle manufacturers in the way their safety systems are implemented. None of these vehicle manufacturers are prepared with their safety implementations to take on the future of autonomous vehicles. The clear gap in preparation can be seen when understanding that there are currently zero systems that have ethics implemented in their safety systems. Having ethics implemented in safety systems has the ability to help more than it does harm, yet due to it being a controversial topic, many car manufacturers and systems developers have elected to ignore the concept. We have the chance to educate these car manufacturers and systems developers to the benefits of implementing ethics into their vehicles. With this, new technologies will be created and accepted that my not pertain to autonomous driving but will use the created systems in new ways. With a clear vision for car manufacturers, systems developers, and future legislature, the autonomous vehicle industry has the possibility of becoming safer for the people who wish to own such vehicles, all while allowing for the creation of new technology that will benefit many more.

There are a few reasons that automated vehicles have grown so rapidly in popularity recently. One of the big reasons is that automated vehicles can be a huge proponent in reducing injuries and deaths caused by car crashes every year. In 94% of cases, crashes are caused by driver error [2].

Driver-, Vehicle-, and Environment-Related Critical Reasons		
	Estimated	
Critical Reason Attributed to	Number	Percentage* ± 95% conf. limits
Drivers	2,046,000	94% ±2.2%
Vehicles	44,000	2% ±0.7%
Environment	52,000	2% ±1.3%
Unknown Critical Reasons	47,000	2% ±1.4%
Total	2,189,000	100%

*Percentages are based on unrounded estimated frequencies
(Data Source: NMVCCS 2005–2007)

Driver-Related Critical Reasons		
	Estimated (Based on 94% of the NMVCCS crashes)	
Critical Reason	Number	Percentage* ± 95% conf. limits
Recognition Error	845,000	41% ±2.2%
Decision Error	684,000	33% ±3.7%
Performance Error	210,000	11% ±2.7%
Non-Performance Error (sleep, etc.)	145,000	7% ±1.0%
Other	162,000	8% ±1.9%
Total	2,046,000	100%

*Percentages are based on unrounded estimated frequencies
(Data Source: NMVCCS 2005–2007)

These crashes are usually due to speeding or recognition error by the driver [2]. Recognition error usually consists of the driver not understanding or noticing the road conditions. Some of these drivers also do not have the ability to understand how their particular vehicle will fair under said road conditions [2].

With automated driving, we hope to remove human error from driving completely. Automated vehicles still take part in crashes but have been the reason for said crashes in only 13% of cases since 2014 [3]. This number is only expected to go down in the future. With improvements in automated vehicles, we have the potential to reduce the risk of injury for everyone who shares the road. But before moving on to improvements that are going to be made, we need to understand where automated vehicles are currently at.

Right now, automated driving systems are in a transition state. They are now starting to be widely accepted, and with that current implementation are trying to rapidly improve in order to be the most appealing to customers. Like I pointed out earlier, a big part of the appeal of automated vehicles is there safety aspect. Currently the safety systems in automated cars contain zero ethics in their implementations [4]. Considering ethics when talking about automated safety systems might seem odd, but we will expand and clarify on this topic later in the paper. For autonomous vehicles right now, a driver must be present and prepared to take control of the vehicle at any moment, meaning they also always need to be paying attention to the road. Automated vehicles are currently at level 3 of automation with level 4.0 expected soon [6]. The levels of automation describe the technical aspect of a system when considering how truly autonomous it is. We are currently at level 3 of automation when considering autonomous vehicles, which means that although the vehicle does not require a person to be monitoring the surroundings at every moment,

it does require a human to be present and paying attention at all times [6]. This is due to the car having no way to diagnose and correct failures in the self-driving system if they were to occur. Level 4 of automation for these systems are coming soon though. Level 4 will require a human to be present simply due to uncertainties outside of the system's control. We have already started to see autonomous vehicles being tested for level 4 automation such as Google's Waymo car.

Automated driving 5.0 is supposed to start appearing in late 2021. This means that vehicles will be able to complete tasks without a human ever needing to be present. With 5.0, some changes need to be made in the infrastructure surrounding automated vehicles. The main change will be that ethics will need to be implemented into the safety system of the car. The best way to describe putting ethics into autonomous vehicles would be the car deciding on how to crash. Right now, when a crash is inevitable and occurring, driving systems are told to not intervene. By intervene I mean that the system should take control enough in order to cause as little damage, or redirected damage, as possible overall to the people inside the vehicle and its surroundings. This can be tough though, because it calls on difficult decisions to be made before implementation. A great example of this is MIT's moral machine [4].

Users are instructed to decide how they would want an autonomous vehicle to crash in a difficult to determine situation. Although the idea might seem tough to digest, implementing studies like these in an autonomous vehicle's system has the ability to do much more good than harm.
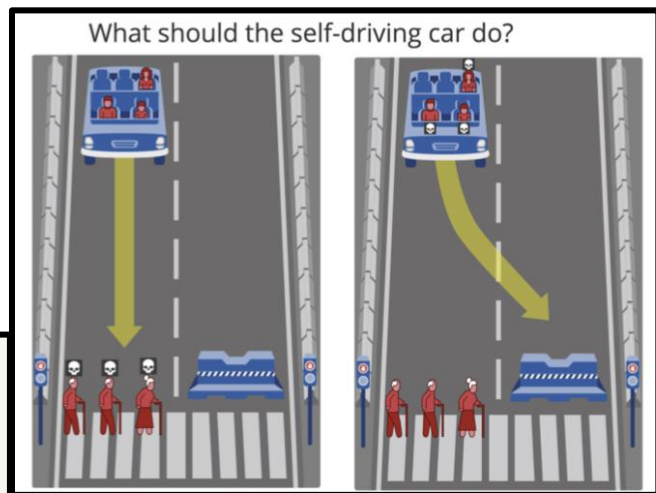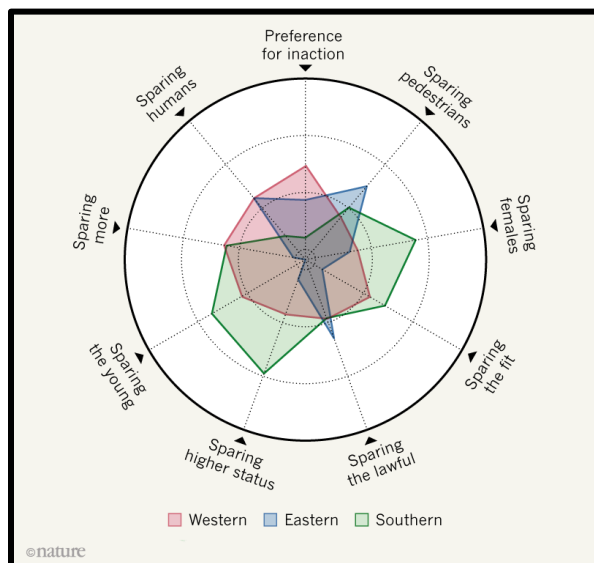
Another layer to the intricacies of ethics in automated vehicles comes from what different societies and cultures believe to be ethical [5].

Determining these ethical standards will come down to the policies and regulations set in place by the companies and governments involved.

There has begun a general movement that companies need to implement ethics into their driving safety systems in the two following ways. The first is that the car needs to determine how to crash, it cannot simply stand idly by as the crash occurs not intervening. The second is that the car needs to determine said way to crash based on inputted factors. These inputted factors will

include where the vehicle will be operated, and the information gathered from the surroundings. This will all be done by the autonomous vehicle's computer, i.e., its AI.

The AI works by taking in information of the vehicles surroundings and making judgements on how the car should operate. Including ethics would help the car minimize damage to its surroundings, most notably in the case of a crash. The vehicle works by taking in data and determining what the next action should be. "A repetitive loop, called Perception Action Cycle, is created when the autonomous vehicle generates data from its surrounding environment and feeds it into the intelligent agent, who in turn makes decisions and enables the autonomous vehicle to perform specific actions in that same environment. [1]" AI should use this data to employ ethics when the vehicle has the chance of crashing. Since this may mean harm comes to others in the hopes of reducing harm to others, the utilitarian approach[3] suits this ethics implementation. Although the utilitarian approach could be used for all autonomous vehicles, the implementation of the ethics should rely on the inputted factors I talked about earlier. The vehicle should decide first how to crash based on the area where the vehicle is operated. If in certain places there is a consensus by the people or law that states the vehicle must crash in a specific way, then the vehicles ethics should be implemented that way. One example of this is that if here in America there is a law that states the people in the vehicle are responsible for the damage that might occur in a crash, that's how the AI should operate. This would be in contrast to another place where there is a law that states the vehicle should minimize the damage done to all variables the vehicle can detect. The second input factor is the surroundings the vehicle can detect. This would be considered when the car is not told specifically by a consensus or law how to crash. If the vehicle is inevitably going to crash and can choose between crashing into a wall or crashing into a field, then the vehicle should choose the field. This becomes increasingly difficult to determine when lives are put into the equation. Should the vehicle alter how it is going to crash if it ends up hurting less people? The utilitarian approach would believe this to be true. If the car can change how it crashes to hurt less people, then it should. Although the car is directly responsible for those people being hurt, it is also responsible for the other people being saved. Autonomous vehicles should implement ethics into their safety systems using a utilitarian approach after considering operation location and taking in information about the surroundings.

The implications of this approach to automated vehicle's AI are that the vehicle will ultimately decide the factors that can be manipulated during the crash and the driver will have no say. If the driver wants to preserve themselves rather than their surroundings, they will not get the choice. Although this seems extreme at first, it makes sense due to the vehicle reducing as much harm caused as possible. This might have an impact in the future that manufacturers can market their cars based on the AI used for their safety systems. If one manufacturer implements an AI that preserves the driver, some people may gravitate that manufacturer's vehicles. This is why I would hope that laws/regulations would come into consideration and force a standard for the ethics implemented in the AI. I could see people debating that the manufacturer is choosing who to harm, and thus makes the implementation of ethics morally prohibited. I think that this could be easily countered though by saying that having the vehicle have the capability of choosing between decisions and not allowing it to happen would be in itself morally prohibited.

As a whole, this study would be representative of AI as a whole. AI should cause as little damage to occur as possible when considering all factors, thus, this case is representative. Although, this does bring up the idea of AI taking drastic measures to reduce human's damage, but it is not pertinent or similar enough to speak on now.

To end off, I want to explain that the development of these autonomous systems will not only apply to people driving places. They can be used in a variety of situations later on that will increase the quality of life for all people. People who choose to distance themselves from using autonomous vehicles will be able to drive on overall safer roads. Autonomous vehicles will be able to take over high health risk jobs. With an increase in the quality of automated driving systems, we have the unique chance to see an autonomous future that does not negatively impact the people who wish to separate themselves from the technology driving it.

Sources

[1] Suhasini Gadam. 2019. Artificial Intelligence and Autonomous Vehicles. (January 2019). Retrieved November 14, 2020 from https://medium.com/datadriveninvestor/artificial-intelligence-and-autonomous-vehicles-ae877feb6cd2

[2] Santokh Singh. 2015. Critical Reasons for Crashes Investigated in the National Motor Vehicle Crash Causation Survey. *Traffic Safety Facts* (February 2015). DOI:http://dx.doi.org/https://crashstats.nhtsa.dot.gov/Api/Public/ViewPublication/812115

[3] Santa Clara University. 2015. A Framework for Ethical Decision Making. (August 2015). Retrieved November 14, 2020 from https://www.scu.edu/ethics/ethics-resources/ethical-decision-making/a-framework-for-ethical-decision-making/

[4] Jamie Leventhal. 2018. In a crash, should self-driving cars save passengers or pedestrians? 2 million people weigh in. (October 2018). Retrieved November 14, 2020 from https://www.pbs.org/newshour/science/in-a-crash-should-self-driving-cars-save-passengers-or-pedestrians-2-million-people-weigh-in

[5] Amy Maxmen. 2018. Self-driving car dilemmas reveal that moral choices are not universal. (October 2018). Retrieved November 14, 2020 from https://www.nature.com/articles/d41586-018-07135-0

[6] Matthew Lynberg. 2020. Automated Vehicles for Safety. (June 2020). Retrieved November 14, 2020 from https://www.nhtsa.gov/technology-innovation/automated-vehicles-safety

Read But Not Used

U, Jenn. "The Road to Driverless Cars: 1925 - 2025." *Engineering.com*, www.engineering.com/PLMERP/ArticleID/12665/The-Road-to-Driverless-Cars-1925--2025.aspx.

Litman, Todd. "Autonomous Vehicle Implementation Predictions Implications for Transport Planning." *Victoria Transport Policy Institute*, 24 Mar. 2020, www.vtpi.org/avip.pdf.

Gomes, Lee. "Hidden Obstacles for Google's Self-Driving Cars." *MIT Technology Review*, MIT Technology Review, 2 Apr. 2020, www.technologyreview.com/s/530276/hidden-obstacles-for-googles-self-driving-cars/.

Zaman, Amin. "10 Astonishing Technologies That Power Google's Self Driving Cars." *National Tyres and Autocare*, www.national.co.uk/tech-powers-google-car/.

Ziegler, Chris. "A Google Self-Driving Car Caused a Crash for the First Time." *The Verge*, The Verge, 29 Feb. 2016, www.theverge.com/2016/2/29/11134344/google-self-driving-car-crash-report.

English, Trevor. "Why We Don't Have Fully Autonomous Vehicles on the Road Yet." *Interesting Engineering*, Interesting Engineering, 9 Feb. 2020, interestingengineering.com/how-do-self-driving-cars-work.

"Gartner Forecasts More Than 740,000 Autonomous-Ready Vehicles to Be Added to Global Market in 2023." *Gartner*, www.gartner.com/en/newsroom/press-releases/2019-11-14-gartner-forecasts-more-than-740000-autonomous-ready-vehicles-to-be-added-to-global-market-in-2023.

Garsten, Ed. "Sharp Growth In Autonomous Car Market Value Predicted But May Be Stalled By Rise In Consumer Fear." *Forbes*, Forbes Magazine, 13 Aug. 2018, www.forbes.com/sites/edgarsten/2018/08/13/sharp-growth-in-autonomous-car-market-value-predicted-but-may-be-stalled-by-rise-in-consumer-fear/#3900c2c5617c.

"Fatality Facts 2018: Yearly Snapshot." *IIHS*, www.iihs.org/topics/fatality-statistics/detail/yearly-snapshot.

"Issues in Autonomous Vehicle Testing and Deployment." *Federation of American Scientists*, 11 Feb. 2020, fas.org/sgp/crs/misc/R45985.pdf.