Ethical Challenges of Autonomous Vehicles

An STS Research Paper submitted to the Department of Engineering and Society

Presented to the Faculty of the School of Engineering and Applied Science

University of Virginia • Charlottesville, Virginia

In Partial Fulfillment of the Requirements for the Degree

Bachelor of Science, School of Engineering

Cooper Dzema

Fall Semester 2021

On my honor as a University Student, I have neither given nor received unauthorized aid on this assignment as defined by the Honor Guidelines for Thesis-Related Assignments

Student

Cooper Dzema Date 11/8/21

Date: 12/02/21

Richard D. Jacques, Ph.D., Department of Engineering and Society

Introduction:

Autonomous vehicles (AVs) will soon be advanced enough to be widely used, but there are numerous still numerous pressing ethical questions such as: Who takes the blame for an AV-related crash? How should AV algorithms decide on the lesser of two evils when a crash is unavoidable? How will data privacy be handled? These questions will need to be answered before widespread adoption of Autonomous Vehicles can take place. Although significant research has been done in the area, there is often disagreement in the literature on what problems should be addressed first as well as how these problems should be addressed. There are many levels of autonomy in AVs, from basic assistance features already in use today to the fully autonomous vehicles of the future. The Society of Automotive Engineers (SAE) defines 5 levels of autonomous vehicles from assisted automation at level 1 to full automation at level 5 (Lim & Taeihagh, 2018). This research paper will focus on some of the key aspects of AV ethics that have been in the centerstage of discussions in literature, and will mostly discuss fully autonomous vehicles (level 5).

The Trolley Problem and Dilemma Situations:

One of the cases that is explored frequently in the literature with respect to AV ethics is the so-called trolley case (Martinho et al., 2020), which was first introduced in 1967 (Etienne, 2020). The simplest trolley case is this; there are 5 people tied down to the tracks in front of an oncoming trolley. You can do nothing and let them die, or pull a lever to divert to trolley to another track that will only kill one person. Is the correct ethical decision to do nothing and let the 5 people die or to pull the lever to save 4 people but in doing so take responsibility for killing the single person on the other track (Fleetwood, 2017).

This case has been studied extensively in the context of AVs in the form of dilemma situations. Though it may not occur often, it is accepted that there will be times when an AV is in a situation where a crash of some sort cannot be avoided, and the AV will have to make a decision as to what action it will take. For example, an AV may be forced to choose between continuing on course and running over a pedestrian or diverting course to crash off the road, potentially injuring the driver. In these situations, the Av must be able to make a choice that is both acceptable to the general public and acceptable to an individual driver. If this case were replaced by one with a human driver, whatever choice made by that individual should not be considered morally wrong, because the human driver only has fractions of a second to react, and their reaction is dictated only by instinct. However, for an AV, this entirely changes. Although an AV would still only have fractions of a second to respond, this would still be plenty of time for the AV's computer to fully understand the situation and be able to make a decision. Additionally, whatever decision the AV ends up making would likely have to be preprogrammed. In short, the decision made by the AV in such a situation is made in anticipation of the situation, and is no longer a result of instinct (Etienne, 2020). Because of this, it becomes important to decide how AVs will react in dilemma situations.

Researchers at MIT developed a study in 2018 called the Moral Machine to collect data on how people would like AVs to react in different situations. The data from this study shows that people have a clear preference for sparing humans over animals, more people over fewer people, and young people over older people. However, the data on whether passengers should be spared over pedestrians or whether taking action is preferable to inaction is much more ambiguous (Awad et al., 2018).

The public will not be comfortable with AVs unless they make ethical decisions, and drivers will not want to use AVs if they are too quick to sacrifice their driver. Research has shown that people would prefer AVs in general to use utilitarian algorithms (ones that do not preferentially protect their drivers and instead seek to minimize overall harm), but would be unlikely to consider purchasing an AV that uses a utilitarian algorithm (Fleetwood, 2017). This dichotomy between how individuals would like AVs in general to behave and how they would want their own AV to behave is a major challenge facing adoption of AVs. If regulations require utilitarian algorithms to be used, then this may delay the public's adoption of AVs. Even though dilemma situations where this would come into play will be very rare, people will likely not trust a vehicle that will not preferentially protect them at all costs. Since it is generally agreed that fully fledged AVs will be safer than human drivers, this would have an overall negative effect on public health. However, if both utilitarian and non-utilitarian vehicles are allowed, then the public would almost exclusively purchase the non-utilitarian models, even though public opinion is that utilitarian models should be used (Bonnefon et al., 2016).

An issue, however, with an algorithm that purely seeks to minimize loss of life is that the people who die due to an AV crash are not always the same people that would die due to a conventional vehicle crash. The vast majority of conventional vehicle crashes are due to human error. In a fully autonomous vehicle crash, the driver is completely innocent, but is still at risk of the AV algorithm deciding to prioritize the lives of others. The result of this is that the victims of AV crashes will more often be innocent than the victims of crashes involving nonautonomous vehicles (Etienne, 2020). Arguments like this must be taken into account when considering whether a purely utilitarian algorithm should be used.

Although the trolley problem and dilemma situations clearly are an issue that will need to be addressed, other scholars contest that this should not be the focus of as much research as it currently is (Himmelreich, 2018; Etienne, 2020; Cuneen et al., 2019). One argument against the widespread usage of the trolley case in AV ethics is that current technology is far from being able to make the types of decisions that studies like the Moral Machine Experiment ask (Cuneen et al., 2019). Additionally, many of these questions should already be morally invalid, such as asking whether an AV should preferentially spare a younger person over an older person. It is clear that an AV algorithm advanced enough to recognize attributes such as age should not be able to use this information to make decisions in a dilemma situation. This is outlined in rule 9 of Germany's Federal Ministry of Transport's report on AVs which states: "In the event of unavoidable accident situations, any distinction based on personal features (age, gender, physical or mental constitution) is strictly prohibited (Germany's Federal Ministry of Transport and Digital Infrastructure, 2017).

Another argument against the focus on dilemma situations is the extreme rarity of their occurrence. Even human drivers do not often encounter a situation where a crash is unavoidable and there is no possible favorable outcome. AVs will encounter these situations even less frequently, due to their superior safety and ability to react quickly. In one of their reports, Intel claims that their AV will always be able to brake before hitting a pedestrian unless they are moving super-humanly fast (Martinho et al., 2020). Although this is clearly a very ambitious goal, and there may still be ways for such a vehicle to encounter a dilemma situation, it is possible that dilemma situations will become so rare that they almost never occur. If this is that case, all of the time and effort spent on looking at dilemma situations could be better applied to more pressing issues such as responsibility or privacy.

Responsibility:

Another ethical issue regarding AVs that has yet to be resolved is the issue of responsibility for crashes. While there is usually a clear culprit in an accident when no AVs are involved, if an AV happens to initiate a crash, it becomes much less clear. Clearly, if the vehicle that causes a crash is fully autonomous, the driver of that vehicle is not responsible, as they had nothing to do with the maneuvers that put the vehicle into that scenario. Although some progress has been made on this issue, no definitive answer has been agreed upon, and many different approaches have been used around the world (Taeihagh & Lim, 2019). Germany's ethics commission on AVs, for example, has stated that responsibility for AV related crashes should shift from the driver to the manufacturer (Germany's Federal Ministry of Transport and Digital Infrastructure, 2017).

For the most part, holding the manufacturer of the vehicle responsible for an AV related crash seems like a plausible solution. This would likely encourage more people to be comfortable driving an AV, as they would not have to worry about liability should the software make a mistake. However, it has also been argued that holding manufacturers responsible like this may discourage innovation, which would have a negative overall effect. This argument states that if manufacturers will be held liable for crashes, this may disincentivize the development of AVs because of the added cost of paying for damages should crashes occur. Since AVs will likely be safer than human drivers, any delay in the adoption of AVs would therefore be a negative outcome (Hevelke & Nida-Rümelin, 2015).

Hevelke & Nida-Rümelin (who conducted the research cited above) concluded that there are 2 ethically acceptable solutions to this issue. The first is to hold the driver responsible if there was time enough to intervene and prevent a crash. This would only be applicable to AVs

that are not fully autonomous. For example, current technologies that allow a car to stay in its lane autonomously still require a human driver to have their hands on the wheel ready to take control at any time. As AV technology advances, however, there will be a point where it is no longer reasonable or even possible for the human driver to intervene. Beyond this point, the driver can no longer be held responsible. The second solution they proposed applies to fully autonomous vehicles and is to hold AV owners collectively responsible for taking the risk of driving an AV. This would likely be accomplished through a mandatory insurance paid by all AV owners (Hevelke & Nida-Rümelin, 2015). This money would be used to resolve crash liability without holding a specific driver responsible.

However, opponents of this stance argue that simply holding manufacturers responsible will not have a significant enough impact to warrant any of these measures (Martinho et al., 2020). As mentioned before, some companies have the goal of creating an AV that is so reliable that it will never be involved in a crash. If most manufacturers can come even close to this level of safety, then the cost of paying for damages would be negligible to them and would not affect them in any meaningful way.

Privacy:

Another ethical concern regarding AVs is data privacy. While non-autonomous vehicles do not need to collect any data in order to function, AVs work by collecting as much data as possible about the surroundings so that their computer can make the best possible decisions. Although this data could in theory only be used by the vehicle it is collected by and then discarded, it has been shown that there are benefits of sharing this data with other AVs. As AVs eventually become common, ability to communicate between separate vehicles could have major benefits. For example, if multiple AVs are on a road and there is an obstacle that is only seen by

the leading vehicle, simulations have shown that the ability of the leading AV to communicate the existence of this obstacle to the vehicles behind it will drastically improve their ability to avoid it. (Cuneen et al., 2019). However, there will need to be a balance between sharing data in order to improve roadway safety and limiting data sharing to preserve privacy (Lim & Taeihagh, 2018).

AVs present such a large privacy risk because they are capable of collecting sensitive personal data such as the exact location of the user and everywhere they go. If not properly regulated, this information can be used in many ways that can negatively affect the AV owner. For example, location data could find its way into the hands of advertisers who could use it to target marketing based on places people frequently visit (Collingwood, 2017). Insurance and credit rating companies could also use data collected by AVs to calculate insurance premiums and credit scores of people based on their driving habits. This method has been shown to be highly inaccurate, and its usage would hurt the consumer (Lim & Taeihagh, 2018).

A common legislative response to the privacy issues of AVs, at least in the United States, is to require AV companies to clearly state exactly what data is collected and how it is used (Lim & Taeihagh, 2018). However, a response like this does not fix the inherent problem. It allows users to make an informed decision about whether they feel comfortable with the information collected, but if all AV companies collect similar data, then people will be forced to either accept the privacy concerns or not use an AV at all. The ownership of data collected by AVs is also still unclear (Collingwood, 2017). If the information collected does not belong to the owner of the AV, then the manufacturer would have every right to use the highly sensitive personal data collected however they choose. If user privacy is to be prioritized, the information collected by AVs should at least be encrypted and anonymous, as any association of data with an individual

AV owner causes numerous privacy concerns. However, it is not clear yet whether this direction will be taken, so the concerns remain.

Conclusion:

This research paper has explored some of the foremost ethical concerns with the emerging technology of autonomous vehicles. There has been much research done in the area, but as can be seen from the literature reviewed, there is still no consensus in many of the most debated areas of AV ethics. Scholars do not agree on the correct approach for AVs to take in dilemma situations, or whether these types of scenarios should even be the focus of research at all. The issue of privacy and the data collection and usage of AVs is something that will soon become a widespread problem if not addressed, but it has been much less prominent in the literature than discussions of dilemma situations, which in reality will occur very infrequently (Martinho et al., 2020). Even simple situations like approaching a crosswalk can raise ethical questions about how an AV should behave (Himmelreich, 2018), but these more mundane questions have received far less attention. A study which examined 238 papers on the subject of AV ethics found that more than half referenced the trolley problem at some point (Martinho et al., 2020). Although the discussion of responsibility for AV crashes has reached more consensus than other areas (it seems mostly agreed upon from the sources examined that the manufacturer should be held responsible), there are still arguments for alternative viewpoints such as the idea of holding all AV owners collectively responsible suggested by Hevelke and Nida-Rümelin.

This is not to say that the introduction of AVs will only cause problems. The one thing that is generally agreed upon in the literature is that AVs will eventually be significantly safer than human driver, and will cause a decrease in accidents. AVs have many other possible benefits, such as increasing the mobility of those who cannot operate a nonautonomous vehicle

and reducing fuel consumption by operating more efficiently than a human driver (Milakis et al., 2017). Although the ethical questions discussed are challenging, if they can be answered, then AV technology could have an overall extremely positive effect.

Reference List:

Awad, E., Dsouza, S., Kim, R., Schulz, J., Henrich, J., Shariff, A., Bonnefon, J.-F., & Rahwan, I. (2018). The Moral Machine experiment. *Nature*, *563*(7729), 59–64. https://doi.org/10.1038/s41586-018-0637-6

Bonnefon, J.-F., Shariff, A., & Rahwan, I. (2016). The social dilemma of autonomous vehicles. *Science*, *352*(6293), 1573–1576. https://doi.org/10.1126/science.aaf2654

Collingwood, L. (2017). Privacy implications and liability issues of autonomous vehicles. *Information & Communications Technology Law*, *26*(1), 32–45. https://doi.org/10.1080/13600834.2017.1269871

Crayton, T. J., & Meier, B. M. (2017). Autonomous vehicles: Developing a public health research agenda to frame the future of transportation policy. *Journal of Transport & Health*, *6*, 245–252. https://doi.org/10.1016/j.jth.2017.04.004

Cunneen, M., Mullins, M., Murphy, F., Shannon, D., Furxhi, I., & Ryan, C. (2020). Autonomous Vehicles and Avoiding the Trolley (Dilemma): Vehicle Perception, Classification, and the Challenges of Framing Decision Ethics. *Cybernetics and Systems*, *51*(1), 59–80. https://doi.org/10.1080/01969722.2019.1660541

Etienne, H. (2020). When AI Ethics Goes Astray: A Case Study of Autonomous Vehicles. *Social Science Computer Review*, 0894439320906508. https://doi.org/10.1177/0894439320906508

Fleetwood, J. (2017). Public Health, Ethics, and Autonomous Vehicles. *American Journal of Public Health*, *107*(4), 532–537. https://doi.org/10.2105/AJPH.2016.303628

Germany's Federal Ministry of Transport and Digital Infrastructure. (2017) Ethics Commission Report on Automated and Connected Driving June 2017. www.bmvi.de

Hevelke, A., & Nida-Rümelin, J. (2015). Responsibility for Crashes of Autonomous Vehicles: An Ethical Analysis. *Science and Engineering Ethics*, *21*(3), 619–630. https://doi.org/10.1007/s11948-014-9565-5

Himmelreich, J. (2018). Never Mind the Trolley: The Ethics of Autonomous Vehicles in Mundane Situations. *Ethical Theory and Moral Practice*, *21*(3), 669–684. https://doi.org/10.1007/s10677-018-9896-4

Lim, H. S. M., & Taeihagh, A. (2018). Autonomous Vehicles for Smart and Sustainable Cities: An In-Depth Exploration of Privacy and Cybersecurity Implications. *Energies*, *11*(5), 1062. https://doi.org/10.3390/en11051062

Martinho, A., Herber, N., Kroesen, M., & Chorus, C. (2021). Ethical issues in focus by the autonomous vehicles industry. *Transport Reviews*, *41*(5), 556–577. https://doi.org/10.1080/01441647.2020.1862355

Milakis, D., Arem, B. van, & Wee, B. van. (2017). Policy and society related implications of automated driving: A review of literature and directions for future research. *Journal of Intelligent Transportation Systems*, *21*(4), 324–348. https://doi.org/10.1080/15472450.2017.1291351

Taeihagh, A., & Lim, H. S. M. (2019). Governing autonomous vehicles: Emerging responses for safety, liability, privacy, cybersecurity, and industry risks. *Transport Reviews*, *39*(1), 103–128. https://doi.org/10.1080/01441647.2018.1494640

Yuen, K. F., Wong, Y. D., Ma, F., & Wang, X. (2020). The determinants of public acceptance of autonomous vehicles: An innovation diffusion perspective. *Journal of Cleaner Production*, *270*, 121904. https://doi.org/10.1016/j.jclepro.2020.121904