Navigating the Road Ahead: Ethical Considerations in Autonomous Vehicle Development

STS Research Paper
Presented to the Faculty of the
School of Engineering and Applied Science
University of Virginia

By

Caroline Peterson

May 10, 2024

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.

ADVISOR

Benjamin J. Laugelli, Assistant Professor, Department of Engineering and Society

Introduction

Ethical concerns arise at the crossroads of innovation and safety in the rapidly evolving landscape of automotive driving systems (ADS) technology. In particular, concerns surrounding the distribution of benefits, the consequences of accidents, and the principles of accountability introduce a new layer of complexity to this landscape. The devastating incident involving an Uber self-driving car on N. Mill Avenue in Tempe, Arizona, which tragically resulted in the loss of a pedestrian's life, has raised questions about the safety and complexities inherent in autonomous technologies. In March 2018, a modified 2017 Volvo XC90 sport utility vehicle (SUV), operating as part of Uber's Advanced Technology Group, was involved in a fatal collision with Elaine Herzberg. An investigation ensued, where the National Transportation Safety Board (NTSB) emphasized safety failures, including lapses by the vehicle operator, Rafaela Vasquez, and lax corporate governance by Uber (NTSB, 2019). The car could not recognize Herzberg, misclassifying her as various objects and failing to slow down to avoid the collision. The NTSB also criticized regulatory oversight of autonomous vehicles as too lenient (NTSB, 2019). The fatal accident marked the first of its kind, setting a precedent for future case studies.

With more analysis of the collision, the validity of ADS implementation is questioned due to the discrepancies between the technology's intended safety advantages and the performance in critical scenarios. In this paper, I will demonstrate that this accident resulted from poor ethical decisions and investigate how Vasquez, Uber, and local governments are culpable for the fatality. Without assessing the moral and ethical responsibilities that automated driving systems hold to the public, I will miss the opportunity to mitigate the fatal flaws within the technology. To accomplish this goal, I will be reviewing multiple official documents, such as the

NTSB report, as well as ADS and law reviews, to reveal the unethical personal and business practices that yielded the accident. Analyzing the case study through the ethical lens of utilitarianism will provide a framework for understanding the ethical dimension of this incident.

Utilitarianism is an ethical framework that provides a theoretical foundation for evaluating the effects or consequences of stakeholders' actions to assess the morality of the situation. The choices of the collaborators and the resulting consequences lend themselves to the use of utilitarianism. I will be reviewing multiple scholarly sources to demonstrate that the development of such technologies does not prioritize safety. More specifically, this investigation will reveal that the stakeholders failed to utilize a practical utilitarian argument, evident in a lack of human agency and misguided accountability. This paper serves as a call to action for society to reevaluate its approach to autonomous vehicles and prioritize all road users' safety.

Background

In 1966, with the formation of the NHTSA and the NTSB, the automotive industry took an important step toward enhancing regulations and prioritizing consumer safety (Mashaw, 2017). In contemporary society, the deployment of autonomous vehicles has required new NHTSA/NTSB safety regulations and, therefore, yielded new classifications for the different stages of automation. According to the Society of Automotive Engineers (SAE) (2016) automation taxonomy guidelines, Automated Driving Systems (ADS) and Advanced Driver Assistance Systems (ADAS) are distinguished by the vehicle's level of autonomy. ADAS are defined as "Level 1 - Driver Assistance" and "Level 2 - Partial Driving Automation" (2016). Conversely, ADS are more evolved and comprise Levels 3-5, which "includes hardware and software that are collectively capable of performing the entire dynamic driving task on a

sustained basis, regardless of [...] the presence of a safety operator" (NHTSA Order No. 2021-01, 2023, p. 8). It is imperative to differentiate the two systems as ADAS mandates that a "human driver must remain fully and continuously engaged in the driving" (NHTSA Order No. 2021-01, 2023, p. 4). Throughout this analysis, ADS will be the primary focus.

Literature Review

The implications of ADS are twofold. On one hand, it heralds a promising era of enhanced safety, with the potential to reduce accidents caused by human error. Jack Stilgoe (2018), a professor at University College London, notes distractions, impairment, fatigue, and speeding - common among human drivers - as major accident contributors. He points to ADS as a solution, potentially saving lives amidst over a million annual global road accident deaths. On the other hand, the Uber crash accentuates the need for robust ethical, legal, and policy oversight to ensure safe navigation in unforeseeable circumstances. Investigating the safety of automated driving systems is critical for individuals and society. Many complex questions need to be answered to ensure all road users' safety while reaping the benefits of ADS.

In considering the moral implications, there is a recognition that moral algorithms for autonomous vehicles pose a social dilemma. While authors like Stilgoe advocate that autonomous vehicles are a popular and effective method to safeguard lives, ethicists Bonnefon, Shariff, and Rahwan suggest differently. They surveyed individuals on the preferred ethics of self-driving cars, and the results revealed that consumers were less likely to purchase an autonomous vehicle that minimized casualties (2016). According to the research, people prioritize their safety and prefer cars that offer protection to passengers, regardless of other factors. It is worth noting that minimizing casualties is the most utilitarian approach.

Despite the personal incentives consumers express, utilitarian ethics ignores the relationships and special moral obligations between individuals. In accidents, the question emerges: Do I have an ethical responsibility to protect individuals inside the vehicle more than outside, potentially contradicting utilitarian principles? Manufacturers and regulators face the challenge of designing algorithms that balance moral values and personal self-interest across different cultures. The ethical considerations of autonomous vehicles stress the need to examine human agency and interpersonal relationships carefully. The differing viewpoints indicate that further research is needed to explore the practical uses of automated driving systems (ADS), particularly in light of the Uber accident in Tempe, Arizona. Scrutinizing the responses of corporations and Vasquez could shed light on a better understanding of the complex moral issues surrounding autonomous vehicles.

Conceptual Framework

A framework is necessary to effectively show autonomous vehicles' ambiguity and demonstrate the ethical responsibility of policymakers, corporations, and individuals. Science, technology, and society (STS) provide a framework for investigating self-driving vehicles through the lens of utilitarian ethics. As a consequentialist theory, utilitarian ethics posits that actions are morally right if they produce the greatest good for most people (Poel & Royakkers, 2011). Consequentialism is a class of ethical theories holding that actions' consequences are central to moral judgments. Taken within the context of ADS, the utilitarian framework isolates the perceived consequences and makes it easier to judge if the decision-making minimized casualties. As a framework, utilitarianism allows for examining the consequences of actions and the principles of impartiality and agent neutrality (Driver, 2014).

Utilitarian ethics assess the consequences of actions by breaking down the essential components of decision-making to determine the influence of their role on the outcome. Jeremy Bentham pioneered this reasoning by evaluating actions on a moral balance sheet, considering intensity, duration, certainty, propinquity, growth, purity, and extent (Brink, 2022). By weighing the different components of a choice, the effects of the action can be calculated and compared to alternatives. The malfeasance of stakeholders can be estimated based on the extent of their influence in the ADS accident. John Mill expanded on Bentham's ideas and emphasized the importance of internal sanctions, such as guilt and remorse, in regulating actions. He introduced the freedom principle through his essay *On Liberty* (1859), which asserts that individuals are free to pursue their pleasure as long as they do not harm others (Müller, 2020). More so, Henry Sidgwick emphasized the egalitarian principle that all individuals' well-being holds significance, i.e., the welfare of one is no greater than that of another (Shultz, 2023). The principle of impartiality and agent neutrality highlights the need for ethical decision-making processes that are free from bias when considering the interests of all relevant stakeholders in the accident.

Despite its strengths, utilitarianism has faced several criticisms. These criticisms suggest that consequences cannot always be objectively measured, leading to uncertainty and unpredictability in moral decision-making. Moreover, utilitarianism may prioritize the happiness of the majority at the expense of minority interests, potentially disregarding individual rights and liberties. The moral ambiguity surrounding the framework prompts the question: Will working toward the betterment of society result in the most favorable outcome for everyone involved? Does the uncertainty of future outcomes render this theory of judgment ineffective? The analysis will delve into these questions to provide a more nuanced approach to decision-making.

Analysis

Rafaela Vasquez

To fully understand and assess the accident, it is imperative to scrutinize Rafaela Vasquez's actions preceding the event. By dissecting her decision-making process, I can reveal the moral shortcomings that played a pivotal role. An important factor in demonstrating the unethical behavior of Vasquez is her human agency. *Human agency* is a moral responsibility that refers to the ability of individuals to make choices that impact the well-being of society (Schlosser, 2019). When considering human agency in the context of utilitarianism, individuals are expected to weigh the potential outcomes of different courses of action and choose the one that produces the most favorable results (Driver, 2014). Take, for example, Rafaela Vasquez and her choices leading up to the accident. While operating as the custodian for Uber's self-driving vehicles, Vasquez decided to shirk her professional responsibility as the "designated driver" and instead chose to watch the reality entertainment The Voice moments before impact (DeArman, 2019). Vasquez's decision to pay attention to entertainment rather than her professional responsibility shows poor human agency. By delving into Vasquez's choices and behaviors, I can uncover the moral failings that contributed to the accident and evaluate the implications for the broader discourse surrounding autonomous driving systems. Through this analysis, I aim to shed light on the ethical complexities inherent in the transition to autonomous vehicles and the moral responsibilities of individuals accompanying such technological advancements.

Although operators like Vasquez serve as the "designated driver," they only "take control" periodically when things go awry. Most operators are not required to intervene, except for Vasquez, who needed to do so for 5.6 seconds over a 39-minute shift (NTSB, 2019). Understanding Mill's version of utilitarian ethics is important because it emphasizes the role of

internal sanctions like guilt and remorse in regulating our actions, which are crucial for moral decision-making (Brink, 2022). When one perceives oneself to be the agent of harm, the negative emotions are centered on the self, leading to feelings of guilt and remorse. When Vasquez is physically driving a vehicle, she can feel the direct consequences of distracted driving. However, with self-driving cars, human agency in decision-making is significantly reduced or even eliminated, which can affect the effectiveness of internal sanctions and individuals' moral repercussions.

The transition to self-driving cars has raised ethical questions about who bears responsibility for the actions and decisions made by these autonomous vehicles. The Arizona Law Review has suggested that if drivers or riders of semi-autonomous vehicles fail to be alert and place too much trust in their cars, they may put themselves in situations where the quick decisions necessary to prevent accidents become impossible (DeArman, 2019). In this specific case, Vasquez prioritized her entertainment over her duty, thus failing to consider the potential consequences of her actions on the well-being of others. Legally, Vasquez's job was to "sit by and take control of the car if necessary" (Diamantis, 2021, p. 10); however, she disputes this. Not only had she driven the car without incident for nine months leading up to the accident, but she had also driven the route where the accident occurred, N. Mill Avenue, 73 times prior (NTSB, 2019). Utilitarianism would dictate that Vasquez should have recognized the potential risks associated with distracted driving and acted to minimize harm and maximize safety. However, her choice to neglect her responsibilities contradicts the principles of utilitarianism, resulting in a tragic outcome.

Jack Stilgoe asserts that "technologies distribute risks and benefits unevenly; they create winners and losers," suggesting that the benefits of ADS and advanced safety features may favor

those who can afford them, potentially creating disparities in access and accountability (2018, p. 55). For example, if corporations can afford to pay settlements and thereby avoid culpability for accidents, while those who cannot afford the technology are held responsible, this could lead to a situation where the benefits of these technologies are distributed unjustly. Indeed, the National Transportation Safety Board (NTSB) board cited Uber's inadequacies in vehicle operator oversight, the pedestrian's unsafe crossing, and the actions of Vasquez as the primary cause of death. As a result, Rafaela Vasquez pleaded guilty to one count of endangerment with three years of probation and no prison time despite initially being charged with negligent homicide (DeArman, 2019). On the other hand, Uber was not found liable and was tasked with implementing a "safety management system" (NTSB, 2019). Vasquez was a scapegoat for Uber and faced the brunt of legal liability despite working for a major corporation that issued the vehicle. Despite this, when a human driver is behind the wheel, they are responsible for making decisions that affect their safety and the safety of others on the road. Vasquez was partly responsible for the accident, as her decision to engage in distractions while operating the vehicle was highly unethical. Vasquez had the free will to make choices that could have prevented the tragic outcome, yet she neglected her duty to prioritize safety. The corporate accountability of Uber will be analyzed in the subsequent section as they designed and implemented the vehicle.

Uber

The advent of self-driving cars has diluted individual agency and redistributed accountability to agencies. Patrick Lin (2016), a philosophy professor and director of Ethics and Emerging Sciences at California Polytechnic State University, underscores the weighty responsibility shouldered by programmers and manufacturers tasked with making life-and-death

decisions without the immediate urgency human drivers face in surprise situations. Instead of the driver remaining culpable, the designers, algorithms, and corporations become liable. Nevertheless, those entities are held criminally accountable for their conduct. According to Marchant and Lindor (2012), automated driving systems (ADS) face the challenge of navigating through unpredictable scenarios, which human drivers typically manage. The Uber vehicle involved in this case study utilizes a "Pilot Assist" algorithm (NTSB, 2019) to detect the environment surrounding the vehicle. This technology remains in its preliminary testing stages as no such self-driving vehicle exists on a production level. Self-driving cars are powered by artificially intelligent (AI) algorithms that use pre-programmed rules and data to make decisions. The onus lies with manufacturers to ensure that AI algorithms do not introduce unforeseeable safety risks and vulnerabilities (Xie, 2020). AI-driven technologies, such as image recognition algorithms, introduce unpredictable risks due to variations in their outputs in different environments. For Uber, the vehicle's radar detected the woman 5.6 seconds before impact, first identifying her as an unknown object, then as a vehicle, and finally as a bicycle. These inconsistencies in radar detection ultimately resulted in her death (Johnson, 2020). The lack of control that lies with the introduction of AI makes drivers unaware of what is to come. As Uber becomes the developer and operator of the vehicle, it is their responsibility to ensure the safety and reliability of its ADS.

Existing ADS liability law fails to adequately incentivize safety, as it allows companies to mitigate their liability for crashes by demonstrating comparative negligence on the part of the plaintiff (Wansley, 2021). This loophole creates a situation where corporations and users engage in a blame game, each attempting to shift responsibility for operating a motorized vehicle onto the other party. Accountability becomes muddled in the aftermath of accidents, such as the one

involving Elaine Herzberg. Despite flaws in Uber's algorithm, which contributed to the fatal accident, and criticism from the National Transportation Safety Board (NTSB) regarding Uber's safety culture, prosecutors refrained from pressing charges against Uber or its employees (Wansley, 2021). Despite a lack of legal consequences, the NTSB places accountability on Uber. It becomes the responsibility of the corporation to predict and implement effective countermeasures to control operator disengagement, ie. Vasquez.

The economic dimension of this case sheds light on Uber's unethical involvement in the accident. From an economic standpoint, Uber and the Arizona government cited "economic growth" as an incentive for inviting Uber's ADs program onto Arizona roads. Arizona Governor Doug Ducey stated, "this is about economic development, but it is also about changing the way we live in work" when discussing the executive order he signed in 2015 supporting the operation of self-driving cars (DeArman, 2019, p. 11). The financial incentives for the implementation of a self-driving fleet of vehicles act as an indicator of Uber's priorities; it shows that Uber executives value the potential profits of technology over the safety of their product. This was further corroborated in 2019 when Uber's CEO, Dara Khosrowshahi, sought to retain the self-driving program and raise the possibility of "licensing its technology to outsider companies" (DeArman, 2019, p. 21). Despite the safety risks, I would posit that the prospect of compensating for the proprietary software and data caused the Uber executives to retain the ADS program.

Best Effort?

While it is evident that both Vasquez and Uber made questionable decisions leading up to the accident, it is essential to acknowledge that they could not have foreseen the fatality. An alternative perspective described by anthropologist Madeleine Elish is the concept of a "moral

crumple zone" (2015). Elish denotes that Vasquez was made an example of, and she served as a symbolic figure tasked with bearing the brunt of responsibility, regardless of the amount of control she had over the situation. This notion suggests that Vasquez, as the operator of Uber's self-driving vehicle, was placed in a position where she was expected to absorb fault. Although Vasquez was visually distracted while operating the vehicle, the NTSB ruled that it was Uber's inadequate oversight of the vehicle operators that facilitated the crash (NTSB, 2019). Uber's failure to enforce the safety management system exacerbated the situation, potentially diminishing Vasquez's capacity to intervene effectively. This highlights the systemic issues within Uber's operational framework, where individual operators like Vasquez may have been ill-equipped to fulfill their duties. The argument can also be made that Uber's self-driving vehicle was not speeding and had the right-of-way (Wansley, 2021). Under these circumstances, it may have been difficult for Uber engineers to have prevented the pedestrian's behavior. However, it is crucial to recognize that the accident occurred within the context of developing technology, where unforeseen challenges and limitations are inherent. Vasquez and Uber may have acted immorally in their decision-making, but arguably, they could not have fully anticipated the fatal consequences of their actions.

Conclusion

In conclusion, the analysis illustrates the multifaceted ethical and practical challenges surrounding integrating autonomous vehicles into transportation systems. The Uber self-driving car accident in Tempe, Arizona, is a poignant reminder of the complexities inherent in autonomous technologies and the urgent need for comprehensive ethical, legal, and policy oversight. Through the lens of utilitarian ethics, it becomes evident that the development of

autonomous vehicles does not prioritize safety, accountability, and the greater good. Examining the actions of Vasquez and Uber reveals significant concerns regarding the widespread adoption of autonomous vehicles. Therefore, while acknowledging the potential benefits of this technology, society must prioritize safety, equity, and the well-being of all road users in the ongoing discourse surrounding autonomous vehicles.

Word Count: 3285

References

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

DeArman, A. (2019). The wild, wild west: A case study of self-driving vehicle testing in Arizona. *Arizona Law Review,* 61(1), 1-25. chrome-extension://efaidnbmnnnibpcajpcglclefindmkaj/https://arizonalawreview.org/pdf/61-4/61arizlrev983.pdf

Diamantis, M. E. (2021). Employed algorithms: A labor model of corporate liability for AI. *Research Handbook on Corporate Liability*, 72(797). https://doi.org/10.4337/9781800371286.00034

Driver, J. (2014, September 22). The history of utilitarianism. In E. N. Zalta (Ed.), *Stanford Encyclopedia of Philosophy* (Fall 2014). Stanford University. https://plato.stanford.edu/entries/utilitarianism-history/

Elish, M. C. (2015, July 25). *When your self-driving car crashes, you could still be the one who gets sued.* Quartz. https://qz.com/461905/when-your-self-driving-car-crashes-you-could-still-be-the-one-who-gets-sued

Johnson, D. G. (2020). Will autonomous cars ever be safe enough? *Engineering Ethics: Contemporary and Enduring Debates.* Yale University Press. https://doi.org/10.2307/j.ctv10sm953

Lin, P. (2016). Why ethics matters for autonomous cars. In Maurer, M., Gerdes, J., Lenz, B., Winner, H. (Ed.). *Autonomous Driving*. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-662-48847-8_4

Marchant, G. E., & Lindor, R. A. (2012).The coming collision between autonomous vehicles and the liability system. *Santa Clara Law Review.* 52, 1321-1340. https://digitalcommons.law.scu.edu/lawreview/vol52/iss4/6/

Mashaw, J. L., & Harfst, D. L. (2017). From command and control to collaboration and deference: The transformation of auto safety regulation. *Yale Journal on Regulation*, 34(1). http://dx.doi.org/10.2139/ssrn.2703370

Müller, V. C. (2020, April 30). Ethics of artificial intelligence and robotics. In E. N. Zalta (Ed.), *Stanford Encyclopedia of Philosophy* (Spring 2020). Stanford University. https://plato.stanford.edu/entries/ethics-ai/

National Transportation Safety Board (NTSB). (2019, November). Collision between vehicle controlled by developmental automated driving system and pedestrian, Tempe, Arizona, March 18, 2018. *Highway Accident Report.* https://www.ntsb.gov/investigations/AccidentReports/Reports/HAR1903.pdf

National Highway Traffic Safety Administration second amended standing general order No. 2021-01 (May 2023). https://www.nhtsa.gov/sites/nhtsa.gov/files/2023-04/Second-Amended-SGO-2021-01_2023-04-05_2.pdf

Poel, I. van de, & Royakkers, L. (2011). Normative ethics. *Ethics, Technology, and Engineering: An Introduction* (2nd, pp. 78–88). Wiley-Blackwell.

SAE International. (2016). Taxonomy and definitions for terms related to driving automation systems for on-road motor vehicles J3016. https://www.sae.org/standards/content/j3016_201806/

Schlosser, M. (2019, October 28). Agency. In E. N. Zalta (Ed.), *Stanford Encyclopedia of Philosophy* (Fall 2019). Stanford University. https://plato.stanford.edu/entries/agency/

Schultz, B. (2023, October 2). Henry Sidgwick. In E. N. Zalta (Ed.), *Stanford Encyclopedia of Philosophy* (Fall 2023). Stanford University. https://plato.stanford.edu/entries/sidgwick/

Stilgoe, J. (2018). We need new rules for self-driving cars. *Issues in Science and Technology*, 34(3), 52–57. https://www.jstor.org/stable/26594264

Wansley, M. (2021). The end of accidents. *U.C. Davis Law Review*, 55(1), 269. https://doi.org/chrome-extension://efaidnbmnnnibpcajpcglclefindmkajhttps://lawreview.law.ucdavis.edu/sites/g/files/dgvnsk15026/files/media/documents/55-1_Wansley.pdf

Xie, G., Li, Y., Han, Y., Xie, Y., Zeng, G., & Li, R. (2020). Recent advances and future trends for automotive functional safety design methodologies. *IEEE Transactions on Industrial Informatics,* 16(9), 5629–5642. https://ieeexplore.ieee.org/document/9026820/