Autonomous Vehicles and Moral Decision-Making: Ethical Responsibility in Artificial Intelligence

Nathan K. Dykema

College of Science and Mathematics, Charleston Southern University

CSCI 332: Applied Networking

Mr. Yu-Ju Lin

10/30/25

Abstract

Autonomous vehicles are becoming rapidly more popular and are on track to become one of the most transformative technologies in modern society. With industries investing heavily in the self-driving platform, they stand on promises of increasing safety, fewer accidents, predictive driving patterns, and more efficiency and accessibility. However, alongside these new benefits comes the potential for very complex ethical concerns and challenges: How should a machine make decisions in situations where harm is unavoidable? Unlike traditional vehicles, autonomous systems rely on algorithms to interpret hundreds of data points to make rapid fire decisions, including those that may involve life or death scenarios. This raises important ethical concerns about responsibility, fairness, and the role of human judgement in artificial intelligence. As autonomous vehicles become the more popular option, the challenge of programming moral and ethical decision making into machines will continue to reveal that machines cannot be given full ethical responsibility, since Artificial Intelligence has no capacity to determine value as a human would. It's important that human oversight always remains in this process, not only to standardize frameworks, but also for accountability.

Autonomous Vehicles and Moral Decision-Making: Ethical Responsibility in Artificial Intelligence

Autonomous vehicles operate using a plethora of sensors, machine learning algorithms that attempt to predict the world around them, and real time data processing. These systems rely on technology such as LiDAR, cameras, radar, and GPS to help them "Sense" their surroundings and feed them data to help them navigate properly. The decision-making process is guided by artificial intelligence systems that interpret the incoming stream of data and quickly determines the appropriate actions that need to be taken, such as braking, accelerating, or steering.

While these systems are designed to prioritize safety, they have their limitations. Machine learning models depend heavily on the data that they receive during their training, which then determines that they may not be able to respond correctly in unexpected or complex situations (Goodall, 2014). In rare cases, such as unavoidable collisions, the system may be placed in a position where decisions it makes impact human lives. This is where ethical considerations become central to the discussion.

A common framework that is utilized in ethical decision-making in autonomous vehicles is the Trolley Problem. Traditionally, this mental experiment is designed to question whether it is morally acceptable to sacrifice one person, when it could save a larger group of people. When applied to autonomous vehicles, the dilemma becomes even more apparent: Should a car prioritize the safety of its own passengers, or should it minimize widespread harm, even if that means between swerving to avoid a group of pedestrians, potentially harming its passenger, or continuing forward and risking multiple pedestrian lives. Situations like this sum up the difficulty of embedding moral reasoning into algorithms. Unlike humans, who may rely on

instinct or emotion to make reactive decisions, machines must follow predefined rules or learned behavior. This raises the next question of whose values get to be programmed into these autonomous systems (Bonnefon et al., 2016). Research from the MIT Media Lab through the Moral Machine Experiment found that public opinion on these dilemmas varies significantly across cultures, suggesting that there is no universal agreement on what the "correct" decision should be (Awad et al., 2018). This lack of consensus makes it even more challenging to standardize ethical decision-making in autonomous systems.

Although ethical dilemmas may seem theoretical, there are many examples of real-world examples in autonomous driving that have already occurred. One of the most popular examples involved a sel-driving test vehicle that struck and killed a pedestrian, raising questions about the reliability and safety of these new "experimental" automated systems. *(National Transportation Safety Board [NTSB], 2019). Such incidents have significantly impacted public trust and brought with it the raised importance of accountability for these systems. Organizations such as the National Highway Traffic Safety Administration have already begun creating guidelines and policies to help regulate autonomous vehicles. However, these regulations often focus on safety standards, rather than focusing on the ethical decision-making frameworks that these vehicles come with. As a result, many ethical questions continue to remain unresolved. Public perceptions also play a large role. Studies document that while people generally support the idea of minimizing harm, they are less willing to purchase vehicles that might sacrifice their own safety for others (Bonnefon et al., 2016). This creates a conflict on the producer side, as it creates an abrasive fight between ethical ideas and consumer preferences, further complicating the issue.

To better understand how autonomous vehicles might handle moral dilemma, it is important to know and examine different ethical theories and their weight on decision-making.

Utilitarianism is an ethical theory that focuses on maximizing overall good and minimizing harm. In the context of autonomous vehicles, this would mean that a vehicle would choose the action that results in the least total harm, even if it involves weighing in the sacrifice of its own passengers. While this approach seems like it might be logical, it raises concerns about whether individuals deciding to purchase or ride in these vehicles would accept putting themselves at risk for the greater good (Goodall, 2014). Another theory is Deontology, which emphasizes rules and duties, rather than outcomes. From this perspective, an autonomous vehicle should follow a list of strict rules, such as not intentionally harming individuals. However, this approach may struggle in situations where harm is completely unavoidable, as it does not provide clear instructions on what to do when duties that are specified are now competing. The last theory that carries weight in the conversation of ethics in autonomous vehicles is Virtue ethics, which is an ethical code focusing on moral character and intentions rather than specific actions. This theory, when applied to machines, can be very challenging, as autonomous systems do not possess a consciousness or moral character to base decisions on. However, developers can attempt to design systems that reflect socially accepted virtues, such as fairness and responsibility. The IEEE has emphasized the importance of incorporating ethical considerations into the design of autonomous system, suggesting that a combination of these frameworks would work best and may be necessary to address the complexity of different scenarios (IEEE, 2019).

One of the most significant ethical challenges associated with autonomous vehicles is attempting to determine where responsibility and accountabilities lie. In traditional driving scenarios, responsibility is typically worn by the human driver. However, when an autonomous system is in controls, the lines become blurred quickly. The problem is now the number of names that reside in the hat of responsibility: The software developers who designed the algorithms, the

company that deployed the vehicle, the user or owner of the vehicle, and the regulatory bodies that determine how it can be used and approved it for use. This creates a legal lattice that seems impossibly complicated to solve. According to the National Highway Traffic Safety Administration, current regulations are still evolving, and there is no universally accepted framework for assigning liability in autonomous vehicle accidents (NHTSA, 2020). Without any clear guidelines or accountability, it becomes difficult to ensure ethical behavior and secure any level of public trust. This highlights the need for transparency for the systems being used and well-defined legal standard.

To address the ethical challenges of autonomous vehicles, there must be a multifaceted approach. One potential solution is the development of standardized ethical guidelines that all autonomous systems follow. These guidelines could be established through collaboration between governments, industry leaders, and academic institutions. Companies must also be part of the conversation and clearly communicate how their systems make decisions, allowing for transparency for users. It is important for them to build public trust. Additionally, audits should be performed regularly on AI systems to attempt to identify and mitigate potential biases or ethical concerns. Regulation will also play a pinnacle role. Governments must establish policies that ensure safety, while addressing ethical considerations. This should include defining liability, setting standards for decisive algorithms, and ensuring compliance is enforced. Lastly, human oversight should always be an essential component. While automated systems can handle many tasks, they require direction and responsibility. Having an element of human involvement can help to ensure that decisions align with societally accepted values.

Autonomous vehicles represent a new and exciting pivot towards the next generation of technological innovations. They offer the potential to completely overhaul public transportation

and improve safety for those that use it. However, it is important to see how they introduce extremely complex ethical challenges that should be considered by those that push this innovation, as well as those that write guidelines that direct its use. As this paper has shown, there is no simple solution to the ethical dilemmas that auto manufacturers face. While different ethical frameworks help to give valuable insight into decision making styles, none can offer a simple solution. Real-world incidents require accountability and transparency on liability issues, as well as ensuring those using this new technology feel safe. Ultimately, the responsibility for ethical decision-making in autonomous vehicles must remain with humans. By developing standardized guidelines, improving transparency, and ensuring accountability, society can work together towards a future where technology aligns with ethical values. As autonomous systems continue to evolve and become integrated, addressing these challenges will be essential to building trust and ensuring their safe deployment.

References

Awad, E., Dsouza, S., Kim, R., Schulz, J., Henrich, J., Shariff, A., … Rahwan, I. (2018). The Moral Machine Experiment. Nature, 563(7729), 59–64.

Bonnefon, J. F., Shariff, A., & Rahwan, I. (2016). The social dilemma of autonomous vehicles. Science, 352(6293), 1573–1576.

Goodall, N. J. (2014). Ethical decision making during automated vehicle crashes. Transportation Research Record, 2424(1), 58–65.

Institute of Electrical and Electronics Engineers (IEEE). (2019). Ethically aligned design: A vision for prioritizing human well-being with autonomous systems.

National Highway Traffic Safety Administration (NHTSA). (2020). Automated vehicles for safety.

National Transportation Safety Board (NTSB). (2019). Collision between vehicle controlled by developmental automated driving system and pedestrian.