Yoko Mohr
CST 300 GWAR
June 11, 2026

Ethical Decision-Making in Self-Driving Cars

Self-driving cars are becoming increasingly common in everyday life. In cities such as San Francisco, AI-powered self-driving cars are now frequently seen on public roads. Companies such as Tesla and Waymo continue to expand the use of self-driving technology in transportation. Many people enjoy experiences using self-driving car technology by purchasing Tesla or hopping onto Waymo taxi services. Even people who do not personally own or use these services share public roads with cars loaded with the technology. Adopting self-driving car technology may provide significant benefits, because studies suggest that self-driving vehicles may reduce accidents that are attributable to human error (Di Lillo et al., 2023; Waymo, n.d.). Such technology would therefore provide significant benefits to society overall, not just the owners or users of self-driving cars. However, ethical concerns still remain regarding how AI-powered self-driving cars should respond during unavoidable crashes. In situations where harm cannot be completely avoided, the system may need to decide who should be protected and who may be sacrificed. According to The Moral Machine Experiment, self-driving cars create moral dilemmas because machines may be forced to distribute harm between different groups of people (Awad et al., 2018).

Imagine a self-driving car surrounded by vehicles on all sides when a large heavy object suddenly falls from a truck directly ahead. The car cannot completely avoid a collision and must instantly decide how to respond. Should it continue forward and risk injuring its passengers to protect the school bus behind it? Should it quickly turn and risk hitting another vehicle on the

right or motorcycle on the left? Should it prioritize minimizing overall harm or protecting the people inside the car? Unlike human drivers, whose reactions are often instinctive and emotional, self-driving cars rely on decisions that are programmed into the system in advance. The owner of the car may suffer the consequences of algorithms that were programmed into the car long before the accident occurred. Would consumers choose a car that always attempts to save the greatest number of lives, or one that prioritizes protecting its passengers at all costs? What happens if self-driving systems begin analyzing and assigning value to the characteristics or lives of different individuals involved in a crash; the very kind of judgment the Moral Machine project asks people to make (Moral Machine, n.d.). As self-driving car technology becomes more common, society must determine what ethical principles should guide these split-second decisions.

This issue has become increasingly important as artificial intelligence rapidly advances and self-driving technology becomes more integrated into society. One side argues that self-driving cars should prioritize protecting their passengers because the car's primary responsibility is the safety of the people inside the car. On the other hand, others argue that self-driving cars should minimize overall harm rather than automatically prioritizing their passengers. Accidents can still happen, and when they do, the decisions are effectively made in advance by programmers, companies, and policymakers. It is easy to simply state that self-driving cars should "minimize harm," but real-world situations can create morally difficult and complicated decisions (Quinn, 2024). Furthermore, who should ultimately make these ethical decisions: programmers, technology companies, or governments?

Stakeholder Analysis

Stakeholders. Two major stakeholder groups exist on opposite sides of the ethical debate surrounding self-driving cars. The first stakeholder group consists of people inside the cars, including drivers and passengers who use self-driving technology. This group is primarily concerned with personal safety, consumer protection, reliability, and trust in the technology. The second stakeholder group consists of people outside the self-driving car, including drivers of nearby vehicles, motorcyclists, pedestrians, cyclists, and other members of the public sharing the roads with self-driving cars. This group is more concerned with public safety, fairness, minimizing overall harm, and protecting vulnerable individuals. Both stakeholders are affected by the ethical decisions programmed into self-driving cars, but each group views the responsibilities of autonomous systems differently.

Stakeholder 1: People inside who use self-driving technology.

Values. People inside self-driving cars value personal safety, reliability, consumer protection, and trust in the technology. Drivers and passengers expect the vehicle to transport them safely and to protect them when danger arises. Because buyers commit a substantial financial investment, their confidence in the system is what drives public acceptance and continued adoption. Reliability matters just as much, since passengers depend on the car to make fast and accurate decisions in an emergency.

Position. People inside the cars argue that self-driving cars should put the safety of their drivers and passengers first when a collision cannot be avoided. In their view, the vehicle's primary responsibility is to the users who depend on it for both transportation and protection. Real-world accidents matter deeply to this group because the outcome of each pre-programmed

decision falls directly on the people inside the car (Shaban, 2025). For that reason, these buyers are increasingly concerned with how autonomous systems will react in a no-win emergency.

Claims. This group supports its position with claims of responsibility, value, and cause. The claim of responsibility holds that the system's first duty is to protect the people using the car. The claim of value is that buyers expect safety and protection in return for what is often a significant financial commitment. Finally, a claim of cause links the two: passenger trust is necessary for widespread adoption, and without that confidence, fewer consumers will embrace the technology; weakening the very public-safety benefits that come from reducing human driving errors, as reported in Tesla's safety data (Tesla, n.d.).

Stakeholder 2: People outside who share the roads with self-driving cars.

Values. People outside self-driving cars value public safety, fairness, the reduction of overall harm, and the protection of vulnerable road users. This group includes drivers of nearby vehicles, motorcyclists, pedestrians, cyclists, and everyone else who shares the road with self-driving cars. Because they are exposed to these systems without ever choosing to use them, they tend to believe an ethical vehicle should weigh the safety of all road users equally rather than automatically favoring the passengers inside.

Position. People outside the cars argue that a self-driving car should minimize total harm in a no-win collision rather than automatically shielding its own passengers. Because these cars operate in shared public space, this group sees them as carrying responsibilities to society as a whole. The decisions encoded in self-driving systems matter to them directly, since the consequences land on nearby drivers, motorcyclists, pedestrians, cyclists, and other people outside the vehicle.

Claims. This group supports its position with claims of fairness, social responsibility, and public safety. The first claim holds that a self-driving car system should prevent the greatest possible number of injuries and deaths when a crash is unavoidable. A second claim of fairness is that public roads affect everyone, not only the owners and passengers of a single car. A third claim extends this to a principle: an ethical system should serve the greatest good for the greatest number rather than automatically privileging the people inside the car.

Argument Question

Should self-driving cars be programmed to prioritize protecting people inside the cars over people outside the car during unavoidable crashes?

Stakeholder Argument

Stakeholder 1: People inside the cars.

Position on Argument. Stakeholder 1 argues that a self-driving car should protect the people inside it first when a collision cannot be avoided. Drivers and passengers depend directly on the cars for both transportation and personal safety. Because consumers choose to invest in this technology, the group holds that the system's primary responsibility is to safeguard its users when danger arises.

Ethical Framework. Ethical Egoism is a moral theory that argues individuals should act according to their own self-interest and well-being. The framework emphasizes self-preservation, self-protection, and rational self-interest rather than sacrificing one's own welfare for others. According to Ethical Egoism, moral decisions should prioritize the interests of the individual or group directly involved in the situation (Moseley, n.d.).

Argument. Applied here, Ethical Egoism suggests that a self-driving car should put its drivers and passengers first, because the people who use the technology trust it to keep them safe. Buyers commit real money to these cars expecting protection in an emergency, and supporters argue that the system's foremost duty is to the safety and well-being of those users when a crash cannot be avoided.

Course of Action. For this group, then, the correct course is to design self-driving cars that protect their drivers and passengers first in a no-win collision. Ethical Egoism backs this approach, since the system's central obligation is to the people who rely on it. At the same time, the group holds that a self-driving car should still work to reduce overall harm wherever doing so does not compromise passenger safety.

Stake. Stakeholder 1 stands to gain or lose a great deal depending on how this question is resolved. If self-driving cars are built to protect their passengers first, the group may benefit from stronger consumer trust, wider adoption of the technology, and improved road safety as human error declines (Tesla, n.d.). If instead these vehicles are programmed to minimize total harm at the passengers' expense, consumers may lose faith in the technology and grow reluctant to adopt it.

Stakeholder 2: People outside the cars.

Position on Argument. Stakeholder 2 argues that a self-driving car should minimize total harm in a no-win collision rather than automatically favoring the people inside it. Because these vehicles operate on shared public roads, the group sees their ethical responsibilities as extending to society as a whole, not only to their passengers.

Ethical Framework. Utilitarianism is a moral theory associated with Jeremy Bentham and John Stuart Mill that seeks to maximize overall happiness and minimize harm for the

greatest number of people. Rather than weighing individual interests, the framework judges actions by their consequences, treating the morally correct choice as the one that produces the greatest good for society as a whole (OpenAI, 2026; Singer, 2026).

Argument. Applied here, Utilitarianism suggests that a self-driving car should prevent as much harm as possible when a crash is unavoidable. Supporters contend that every human life should count equally, rather than the passengers inside automatically taking precedence. Since these self-driving cars operate in public and affect many people at once, the ethical priority is to minimize total injuries and deaths. Buying the technology itself does not grant a passenger greater moral worth than the road users outside the car (OpenAI, 2026).

Course of Action. For this group, the right course is to program self-driving cars to minimize overall harm when a collision cannot be avoided. Utilitarianism supports this, since the goal is the greatest good for the greatest number. In their view, a self-driving car system should judge each situation by its total consequences rather than by the self-interest of its passengers.

Stake. Stakeholder 2 likewise has much at stake in how the question is resolved. If self-driving cars are designed to minimize overall harm, society may benefit from greater public safety, fewer injuries and deaths, and broader protection for everyone on the road. If instead these self-driving cars protect their passengers above all others, the people outside may feel less safe on public roads, and some may lose faith in the fairness of self-driving car systems altogether (OpenAI, 2026).

Student Position

Opinion. Consumers who invest in self-driving technology reasonably expect these systems to prioritize their safety and the safety of their passengers. While self-driving cars hold real promise for improving public road safety, that promise should not come at the expense of the

very people who adopt the technology (Di Lillo et al., 2023; Waymo, n.d.; Tesla, n.d.). My position is that, although self-driving cars may already make roads safer in the general case, they should prioritize protecting the people inside the car during unavoidable crashes.

The broader safety argument is strong. Self-driving cars have the potential to reduce accidents caused by drunk driving, distracted driving, road rage, fatigue, and ordinary human judgment errors. Unlike human drivers, AI systems do not become emotional, reckless, tired, or impaired; they are designed to respond consistently to road conditions and emergency situations. These advantages matter, but they are fully realized only if the public actually trusts and adopts the technology. Greater public trust increases adoption, and increased adoption could reduce traffic-related injuries and deaths across society as a whole. Passenger protection and public safety are therefore not opposites; they reinforce each other.

Alignment. My position aligns most closely with Stakeholder 1: the people inside the cars, the drivers and passengers. Human drivers are inconsistent and often emotionally or physically impaired, which is precisely the problem self-driving car systems are meant to solve. For that solution to reach the road, however, public trust and consumer confidence are essential to the technology's future development. Systems that prioritize passenger safety, an approach rooted in Ethical Egoism, are more likely to gain the widespread, real-world adoption needed to make any difference at scale.

This is where my view bridges two ethical frameworks. Wider adoption of self-driving cars may reduce accidents caused by human error, which means an Ethical Egoist design could still serve one of the central goals of Utilitarianism: reducing overall harm across society. The two are not as opposed as they first appear. At the same time, ethical decision-making in a

real-world crash is far more complicated than a simple mathematical calculation of total harm. Technology development depends on sponsorship, a wide variety of investors, societal involvement, consumer trust, and long-term public acceptance. For all these reasons, ethical programming standards should not rely on overly simplistic or purely mathematical rules. Even so, the two frameworks are not identical: Ethical Egoism protects a specific group, the passengers, while Utilitarianism weighs everyone equally. They can align in outcome without collapsing into the same principle.

Recommendation. Based on this position, I offer several recommendations. First, governments and technology companies should work together to create transparent ethical standards rather than leaving these decisions to any single manufacturer. Second, because a crash unfolds in an instant, a human must make the programming decision in advance, encoding it in software so the self-driving car can act quickly and automatically in unavoidable cases. Third, companies should publicly share testing data, safety reviews, and ethical guidelines to strengthen public trust and accountability (Goodall, 2023). Fourth, self-driving cars should prioritize protecting their passengers while still attempting to minimize overall harm wherever possible; these goals can coexist. Finally, self-driving systems should continue to improve through ongoing testing, oversight, public involvement, and technological advancement, all of which move us toward a safer driving environment for everyone.

References

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

Di Lillo, L., Gode, T., Zhou, X., Atzei, M., Chen, R., & Victor, T. (2023, September 3). *Comparative safety performance of autonomous- and human drivers: A real-world case study of the Waymo One service*. arXiv. https://arxiv.org/abs/2309.01206

Goodall, N. (2023, November 10). *A note on Tesla's revised safety report crash rates*. arXiv. https://arxiv.org/abs/2311.06187

Moral Machine. (n.d.). *Moral machine*. Massachusetts Institute of Technology. https://www.moralmachine.net/

Moseley, A. (n.d.). *Egoism*. Internet Encyclopedia of Philosophy. Retrieved June 1, 2026, from https://iep.utm.edu/egoism/

OpenAI. (2026). *ChatGPT* (June 5 version) [Large Language model]. https://chatgpt.com/

Quinn, T. (2024, September 26). If a car can drive itself, can it make life-or-death decisions? *Harvard Business School Working Knowledge*. https://www.library.hbs.edu/working-knowledge/if-a-car-can-drive-itself-can-it-make-life-or-death-decisions

Shaban, B. (2025, January 20). While Waymo not blamed in multi-car wreck, it's the first fatal collision involving a driverless car. *NBC Bay Area*. https://www.nbcbayarea.com/investigations/waymo-multi-car-wreck-san-francisco-driverless/3766860/

Singer, P. (2026, April 28). *ethics. Encyclopedia Britannica.* https://www.britannica.com/topic/ethics-philosophy

Tesla. (n.d.). *Full self-driving (supervised) vehicle safety report*. Retrieved May 20, 2026, from https://www.tesla.com/fsd/safety

Waymo. (n.d.). *Waymo safety impact*. Retrieved May 20, 2026, from https://waymo.com/safety/impact/