Ethics of Autonomous Vehicles: Standards and Enforcement of Rational Behavior

A 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

Haotian Ren

Spring 2023

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

Joshua Earle, Department of Engineering and Society

Introduction

In General Motor (GM)'s 1939 exhibit, an American theatrical and industrial designer, Norman Bel Geddes, created the first self-driving car. This first autonomous vehicle ever on earth was an electric vehicle guided by radio-controlled electromagnetic fields generated with magnetized metal spikes embedded in the roadway. It sounds like the settings of a magnetic train, or Maglev, but, instead, the magnets were used to guide the vehicle rather than making it float. By 1958, another design was invented by GM. The car's front end was embedded with sensors called pick-up coils that could detect the current flowing through a wire embedded in the road. The current could be manipulated to tell the vehicle to move the steering wheel left or right. This is a significant improvement as sensors were added to vehicles for the first time. In 1977, the Japanese improved upon this idea, using a camera system that relayed data to a computer to process images of the road. However, this vehicle could only travel at speeds below 20 mph. The improvement came from the Germans a decade later in the form of the VaMoRs, a vehicle outfitted with cameras that could drive itself safely at 56 mph. (*History of the autonomous car*)

As technology has improved over the last 30 years, so does self-driving vehicles' ability to detect and react to their environment. Today, car makers and technology giants have started to focus on inventing their autonomous driving system, the "brain" of autonomous vehicles. Tesla, the pioneer in electric vehicles, unveiled their autonomous driving system, Autopilot, in 2015. Nowadays, many companies have mastered their level 3 autonomous driving system, which can drive itself on highways and city roads except during rush hours. This level of automation still requires drivers always to put their hands on the steering wheel in case the system can hardly

deal with emergent situations. Other companies, like Ford, said they are skipping level 3 automation and going straight to level 4. (Faggella, 2020)

While researchers and scientists are devoting their time to the next generation of self-driving cars, concerns about this emerging technology have started occupying the public's view. On August 12, 2021, the founder of a famous Chinese restaurant brand died while driving an NIO ES8 with the NOP function on. NOP is not even a level 3 autonomous driving system, just a driving assistant module. (Tang, 2022) From July 2021 through May 2022, 11 fatal crashes were reported to NHTSA, National Highway Traffic Safety Administration. Of the death, 5 were related to Tesla's autonomous driving system. (Los Angeles Times, 2022) "Cars are not just technological systems; they are sociotechnical systems." (Johnson, 2020) Does this technology really lower the rate of car accidents? How will the safety of passengers and others in traffic with self-driving cars be ensured? Could the autonomous driving system think ethically?

The first two questions will be answered with the improvement of autonomous driving technologies and the enforcement of reasonable tests on cars before their appearance on public roads. However, it's not easy to answer the third one. Even for some proven technologies, there are ethical issues and concerns related to them nowadays—for example, misuse of private information collected through mobile devices. However, as this technology advances at a fast pace, it's time to start thinking about ethical aspects and the social effects it could bring. There are expected to be 3.5 million cars with full automation by 2025 and 4.5 million by 2030. (*Autonomous vehicles*) Of course, there will not be a definite answer to this question. In this paper, the history of automation and ethics will be analyzed first. Then, after introducing official

policies and standards, it will be explored whether autonomous vehicles could ever be ethical, in other words, making rational decisions.

Ethics in Automations

When it comes to topics on ethics, the first thing that comes to our mind should be the famous trolley problem. In the philosophical dilemma, a trolley is out of control and headed toward five people who are on the trolley track and will surely die if the trolley continues on its path. However, there is a switch that changes the tracks, and if one flips the switch, the train will be diverted to another track. Unfortunately, there is one person on the alternative track. Thus, if one flips the switch, five people will be saved, and one will die. The dilemma is whether one should flip the switch. The justification for doing so is utilitarian; although one death is terrible, it is less harmful than five deaths. The problem is that in flipping the switch, one will be intentionally causing the death of the person on the alternative track. If one does nothing, five people will die, but one will not be the cause of their death. This dilemma illustrates the difference between killing someone and letting someone die, as well as between utilitarianism and Kantian theory or any other theory based on intentions. (Johnson, 2020)

Some would argue that the trolley problem is irrelevant to the design of autonomous vehicles. They believe autonomous vehicles will not encounter such a dilemma that requires them to make decisions according to moral permissibility. The only rule the designers should follow is to do their best to let the autonomous vehicle avoid accidents. However, it’s not entirely impossible for an autonomous vehicle to face a trolley problem in practice. For example, suppose that a motorcyclist is skidding across the road toward a crowd of pedestrians on the pavement. The

autonomous vehicle can brake, in which case the motorcyclist will skid into the pedestrians and cause their death. The autonomous vehicle could also accelerate into the motorcyclist, in which case the motorcyclist would be killed, but the skid would be deflected, and the pedestrians would be unharmed.(Keeling, 2019)

These scenarios are far beyond the scope of this paper. There is not yet a unified answer to the trolley problem, and in real life, hardly could anyone really comes into such a dilemma. On the other hand, even for a human being, it is not easy to evaluate which kind of behavior is moral, not to mention an autonomous vehicle with its "brain" invented by humans. Ethics can be viewed as either a rational choice or an intuitive choice (Rhim, 2021). Instead of exploring more complicated intuitive moral decisions, let us first focus on whether autonomous vehicles nowadays can make rational decisions. In the view of rationalists, a decision must have a consistent rationale behind it, or else it is not an ethical decision. (Hooker, 2008) For example, when an autonomous vehicle encounters a pedestrian crossing a road, with enough distance from the pedestrian, the autonomous vehicle should slow down and wait instead of steering the wheel and passing by the pedestrian breathtakingly like it usually happens in an action movie.

Back to The Time Without AI

Why do people start worrying about the ethics of automation related to artificial intelligence, the fourth industrial revolution? (Bai, 2020) Traveling back in time to 100 years ago, right after the second industrial revolution, when advancements in manufacturing and production technology enabled the widespread adoption of technological systems such as telegraph and railroad networks, gas and water supply, and sewage systems, people already started discussing ethics of

the appearance of machines. However, the perspective of ethics they focused on was largely different from that today. "The old partnership between the producer and the consumer is dead. The skillful artificer who loved the offspring of his labour is gone, and his degenerate successor is the salesman whose methods excite no condemnation … the art of getting rid of commodities, that is, of inducing the consumer to buy what he does not want-for, surely no great art is needed to persuade a man to buy something he wants, that he can afford and that is worth the money." (Freeman, 1923) Instead of discussing whether machines were moral, people at that time, when machines that could do mass production first appeared, cared more about the quality of goods produced by machines. For example, the mass use of stamped sheet steel in tools, containers, etc., did not provide a good experience for users, though it was cheaper than the original materials used for those goods.

In contrast, today, people worry more about machines' morality. Why? Regarding the machines back to the second industrial revolution, they were built to accelerate production efficiency. Usually, an input was given by a human, and an output, a product or service, was provided by the machine. Several advanced types of machines have employed feedback loops, for example, the air conditioning system, which can stop automatically when the room temperature reaches the set value. However, these machines all have one common trait: their reactions are hard-coded and predictable. In other words, they cannot think.

The Era of Autonomous Vehicles

In 1950, while working at the University of Manchester, A. M. Turing, the father of theoretical computer science and artificial intelligence, introduced the Turing test. “I propose to consider the

question, ‘Can machines think?” (Turing, 1950) Turing describes the new form of the problem in terms of a three-person game called the “imitation game”, in which an interrogator asks questions of a man and a woman in another room in order to determine the correct sex of the two players. Turing’s new question is: “Are there imaginable digital computers that would do well in the imitation game?”

Apparently, autonomous vehicles, a successful application of artificial intelligence, can imitate human beings in driving a car. When a machine starts to “think” and make unpredictable decisions, there are ethical problems. Hard-coded machines can make rational choices because they are programmed to replicate what humans do. Nevertheless, complicated artificial intelligent machines, though expected to imitate human behavior, may end up executing normal operations in a random sequence, leading to a terrible result, like the fatal crashes of autonomous vehicles mentioned in the introduction.

Though designed with rigorous theories and formulas, neural networks cannot generalize a task requiring implicit logic from physical reasoning. (Traylor, 2022) For a complicated deep neural network, the brain of an autonomous vehicle not even could the designer predict its output. Suppose that the output of such a neural network can drop randomly at any point in an infinitely large space; the easiest way to make it predictable is to bind it within a small cube of fixed length. Safety modules like AEB (Autonomous Emergency Brake) or FCW(Forward Collision Warning) have been invented on the manufacturer's side. (*Forward collision warning (FCW) and automatic emergency braking (AEB)*) On the regulator side, standards and policies might be a solution to force autonomous vehicles to be rational.

Standards & Policies

Autonomous vehicles are expected to increase road safety by reducing the number of accidents and severity of crash consequences by making more rational decisions. (Rhim, 2021) Rarely are there any standards, policies, rules, or guidelines about the safety of autonomous vehicles, but the condition for rational decision-making of autonomous vehicles is not better. In the United States, the NHTSA has issued guidelines for states to use in assessing whether autonomous cars are ready to be put on public roads. However, the guidelines are broad and leave it up to each state to perform its evaluations. In addition to federal guidelines, many states in the United States have passed legislation regarding autonomous vehicles—not all of it related to safety—and several states have granted licenses to auto manufacturing companies to test experimental self-driving cars. (Johnson, 2020) Though ensuring that autonomous vehicles are safe requires more than just rational behaviors, guidelines on the safety of autonomous vehicles contain some perspectives of the authorities on ensuring rational decision makings of autonomous vehicles. The following section will compare the difference in guidelines between the NHTSA and IEEE (Institute of Electrical and Electronics Engineers), one of the top organizations for computer-related science.

The Government vs The Top Institute

Based on the introduction, the guideline from NHTSA is designed for generalized use by the manufacturer of both automated and traditional vehicles. The guideline focuses on a wide range of areas, including data privacy, cybersecurity, registration, etc. The two essential parts for us are system safety and ethical considerations. “Manufacturers and other entities should follow a robust design and validation process based on a systems-engineering approach to design HAV

systems free of unreasonable safety risks." (*Federal Automated Vehicles Policy - september 2016*) Disappointingly and expectedly, there is no helpful information except some qualitative "requirements" that can hardly be verified. However, this government document has some good perspectives in the ethical considerations section. "For example, most States have a law prohibiting motor vehicles from crossing a doubleyellow line in the center of a roadway. When another vehicle on a two-lane road is double-parked or otherwise blocking a vehicle's travel lane, the mobility objective (to move forward toward an intended destination) may come into conflict with safety and legality objectives (e.g., avoiding risk of crash with oncoming car and obeying a law). An HAV confronted with this conflict could resolve it in a few different ways, depending on the decision rules it has been programmed to apply, or even settings applied by a human driver or occupant." It also addresses the dilemma when "the safety of one person may be protected only at the cost of the safety of another person." However, it is like mathematical proof without a conclusion. No guidelines with suggested solutions are given, except for a sentence telling manufacturers to address this problem appropriately.

Quite differently, the document by IEEE focuses more on the quantitative side. Not only are there equations restricting accelerations on the ego vehicle under some specific scenarios, but also are there considerations of osculated road users, including cyclists and pedestrians. Though it does not contain any ethical considerations on automated vehicles, many practical scenarios and expected behaviors of an autonomous ego vehicle can be referenced to test the safety of an autonomous vehicle. For instance, one scenario assumes that "the ego vehicle is traveling along a road with other road users moving in the opposite direction at a non-junction; the ego vehicle's path may temporarily intersect with the other road user's path (e.g., while performing a legal

passing maneuver); a potential front collision is assessed as avoidable; therefore, no emergency maneuver is required." Though the guideline does not mention any dilemma, it gives a solution when a dilemma similar to the one inside the NHTSA guideline happens. "For example, in an emergency situation, the ADS(automated driving system)-operated vehicle may perform a high-g maneuver to avoid a frontal collision, accepting the residual risk of a potential rear-end collision."(*IEEE Standard for Assumptions in Safety-Related Models for Automated Driving Systems*)

Enforcements

Code of ethics, a trendy thing in the society of engineers, "represent the collective wisdom of engineers." It not only guides engineers to behave ethically and help each other but also requires engineers to follow the privacy rule and not share unauthorized data. However, "they are ineffective because they have no enforcement power." (Johnson, 2020) The code of ethics is just a piece of paper without a good way to verify that people are not revealing unauthorized data. So are the rules for autonomous vehicles.

The IEEE guideline document provides some validation methods to test the safety of an autonomous vehicle. "The following systematic process standards may lead to the definition of a safety-by-design architecture, which may then lead to the use of formal verification techniques with a formal model, which can be further analyzed using robustness techniques, which then can be tested within a simulator, on closed courses, and then on public roads." (*IEEE Standard for Assumptions in Safety-Related Models for Automated Driving Systems*) The document by HNTSA also addresses the importance of verification. Waymo, a leading electric vehicle

manufacturer, developed a benchmark called Waypoint to test its autonomous vehicles. (*Waymo's safety methodologies and safety readiness determinations*) Months ago, Mercedes-Benz announced that they acquired a level 3 license from SAE (the Society of Automobile Engineers), which allow them to sell cars with fully autonomous driver assist system in the U.S.(Tarantola, 2023). Despite the fact that Waymo is testing its vehicle with a tool developed by itself, it is more urgent to worry about those car manufacturers who are testing their autonomous vehicles with tests for traditional vehicles.

Some states in the U.S. have legislation related to the safety of autonomous vehicles, forcing car manufacturers to ensure that their autonomous driving systems can behave rationally under several circumstances. However, legislation specific to the safety of autonomous vehicles is still a blank field in 17 states in the U.S. (*Report autonomous vehicles: Self-driving vehicles enacted legislation*) "There seems little doubt that standards, testing, regulations, and legislation will play an important role in ensuring the safety of autonomous cars of the future, but for now it is unclear what approach is the best." (Johnson, 2020)

Conclusion

Where is the future of ethics about autonomous vehicles going? As suggested by NHTSA, "A combination of some of the new regulatory tools (in conjunction with existing tools and authorities) may help to advance the goals of long-term safety regulation and safe deployment of HAVs" (*Federal Automated Vehicles Policy - september 2016*) This includes both certification systems by a third-party authority like SAE, or self-certification systems like Waypoint. "From

the earliest years of the Agency's history, sometimes in response to the Agency and sometimes on its own initiative, Congress has taken action to address these challenges with legislation refreshing and modernizing the Vehicle Safety Act." With potential improvements in these tools and laws, it is possible to ensure the rationality of autonomous vehicles soon. Nevertheless, this is just a tiny part of the ethics of autonomous vehicles.

There is still a long way to go on the road of ethics. Say the famous trolley problem; in a similar scenario where the autonomous vehicle chooses to do an emergency maneuver to avoid death to a road user but causes some injuries to an innocent cyclist, how should the liability be decided? Should the human driver be responsible for the consequence? Should the car manufacturer be responsible for it? Or should the software engineer who programmed the autonomous driving system that made a possibly "unethical" choice go to jail? Social impacts like job displacement are also worth considering, like people's worries after the second industrial revolution.

Nevertheless, the most challenging part behind all of these is moral decisions. There has yet to be an effective method to measure the morality of a human being, not to mention that of artificial intelligence. AI and robotics have raised fundamental questions about what we should do with these systems, what the systems themselves should do, and what risks they have in the long term. They also challenge the human view of humanity as the intelligent and dominant species on Earth. We have seen issues that have been raised and will have to watch technological and social developments closely to catch the new issues early on, develop a philosophical analysis, and learn for traditional problems of philosophy. (Müller, 2020) So far, what can be done is to use quantitative ways to ensure a qualitative behavior of AI. It is important to ensure that the

development and deployment of autonomous vehicles are guided by ethical principles that are safe, reliable, and beneficial to society.

Bibliography

Autonomous vehicles. NAIC. (n.d.). Retrieved April 21, 2023, from https://content.naic.org/cipr-topics/autonomous-vehicles

Bai, C. (2020, May 6). *Industry 4.0 technologies assessment: A sustainability perspective*. International Journal of Production Economics. Retrieved April 21, 2023, from https://www.sciencedirect.com/science/article/pii/S0925527320301559?via%3Dihub

Faggella, D. (2020, March 14). *The self-driving car timeline - predictions from the top 11 global automakers*. Emerj Artificial Intelligence Research. Retrieved April 21, 2023, from https://emerj.com/ai-adoption-timelines/self-driving-car-timeline-themselves-top-11-automakers/

Federal Automated Vehicles Policy - september 2016. U.S. Department of Transportation. (n.d.). Retrieved November 30, 2022, from https://www.transportation.gov/AV/federal-automated-vehicles-policy-september-2016

Forward collision warning (FCW) and automatic emergency braking (AEB): Why they are essential? Highway Code Resources. (2023, January 22). Retrieved April 21, 2023, from https://mocktheorytest.com/resources/forward-collision-warning-fcw-automatic-emergency-braking-aeb

Freeman, R. A. (1923, July). *Some ethical consequences of the industrial revolution - JSTOR*. Retrieved April 21, 2023, from https://www.jstor.org/stable/2377592

History of the autonomous car. TitleMax. (2020, April 30). Retrieved April 21, 2023, from https://www.titlemax.com/resources/history-of-the-autonomous-car/

Hooker, J. (2008, February). *Ethics as rational choice*. Retrieved April 21, 2023, from http://public.tepper.cmu.edu/ethics/rationalchoice.pdf

IEEE Standard for Assumptions in Safety-Related Models for Automated Driving Systems, in IEEE Std 2846-2022 , vol., no., pp.1-59, 22 April 2022, doi: 10.1109/IEEESTD.2022.9761121.

Johnson, D. G. (2020). *Engineering ethics: Contemporary and enduring debates*. Yale University Press.

Keeling, G. (2019, March 4). *Why trolley problems matter for the ethics of Automated Vehicles - Science and Engineering Ethics*. SpringerLink. Retrieved April 21, 2023, from https://link.springer.com/article/10.1007/s11948-019-00096-1

Los Angeles Times. (2022, October 19). *11 more crash deaths are linked to automated-tech vehicles*. Los Angeles Times. Retrieved April 21, 2023, from

https://www.latimes.com/business/story/2022-10-18/11-more-crash-deaths-are-linked-to-automated-tech-vehicles

Müller, V. C. (2020, April 30). *Ethics of Artificial Intelligence and Robotics*. Stanford Encyclopedia of Philosophy. Retrieved April 21, 2023, from https://plato.stanford.edu/entries/ethics-ai/#Clos

Report autonomous vehicles: Self-driving vehicles enacted legislation. National Conference of State Legislatures. (n.d.). Retrieved April 21, 2023, from https://www.ncsl.org/transportation/autonomous-vehicles

Rhim, J., Lee, J.-H., Chen, M., & Lim, A. (2021, April 7). *A deeper look at Autonomous Vehicle Ethics: An Integrative Ethical Decision-making framework to explain moral pluralism*. Frontiers. Retrieved April 21, 2023, from https://www.frontiersin.org/articles/10.3389/frobt.2021.632394/full

Tang, Y. (2022, May 29). *Entrepreneur died in NIO ES8 Crash 'trusted' nop so much he used it for 52% of his July driving miles*. CnEVPost. Retrieved April 21, 2023, from https://cnevpost.com/2021/08/17/entrepreneur-died-in-es8-crash-trusted-nop-so-much-he-used-it-for-52-of-his-july-driving-miles

Tarantola, A. (2023, January 27). *Mercedes is the first certified level-3-autonomy car company in the US*. Engadget. Retrieved April 21, 2023, from https://www.engadget.com/mercedes-first-certified-level-3-autonomy-car-company-us-201021118.html

Traylor, A., Feiman, R., & Pavlick, E. (2022, September 29). *Can neural networks learn implicit logic from physical reasoning?* OpenReview. Retrieved April 21, 2023, from https://openreview.net/forum?id=HVoJCRLByVk

Turing, A. M. (1950, September). *COMPUTING MACHINERY AND INTELLIGENCE*. Academic.oup.com. Retrieved April 21, 2023, from https://academic.oup.com/mind/article/LIX/236/433/986238

Waymo's safety methodologies and safety readiness determinations. (n.d.). https://storage.googleapis.com/sdc-prod/v1/safety-report/Waymo-Safety-Methodologies-and-Readiness-Determinations.pdf