John-Nicholas Krinos
ENC 2135 – Fall 2024
Prof. Bronstein
Project 1 Essay – Final Draft
October 8th, 2024

# The Life and Death of Self-Driving Car Algorithms

Technology controls daily life in the modern day, making programming ethics more complicated than ever. Artificial Intelligence (AI) has become increasingly advanced in recent years, processing more data through the support of huge data centers worldwide. Self-driving vehicles (SDVs) use AI to operate via neural networks and machine learning, using existing data and statistics to identify patterns that determine what the car should do. After executing the suggested action, they get feedback from the user or a trainer program on whether their answer is accurate. Programmers can adjust these training programs or give their testers specific tasks that change the model's goals and subsequent development. They become complex over time and perform different actions based on the input and response training sessions. AI algorithms can solve problems without encountering the conflicts humans encounter when solving them, presenting friction between human values and computer programming. AI and automobiles are pressing examples: SDVs now being tested worldwide highlight many moral and ethical contradictions that can make their programming difficult to justify.

## Evolving Artificial Intelligence

To discuss SDVs, we must begin with the invention of the computer. According to Ophthalmology professor Andrzej Grzybowski, in about 1822 the first computer was made by Charles Babbage and his colleagues who based it on a "Jacquard loom", a machine used to methodically sew garments used before the invention of sewing machines (Grzybowski 1). From there, the next major event was in 1950, when Alan Turing proposed the "Turing Test", the first

intelligence test for a computer (Grzybowski, Para. 19). These computers were much less complicated than modern AI:  they were tested solely by having a conversation with a human without the human realizing that it was speaking to a computer. As mentioned by electrical engineering graduate Nikesh Muthukrishnan, at around the same time, McCulloch and Pitts published a computer model that could function "comparable with neurons in the human brain" (Muthukrishnan Para. 3). The first recorded use of the term "Artificial Intelligence" in the context of a computer was by John McCarthy in 1956, and since then the term has become popular, more recently being used in many publications and articles as a hook word to draw people in as AI has become more widely recognized and respected (Grzybowski Para. 26). Later in the 1980s, an explosion occurred in the development of AI, and some of the models we use today came out, such as the "deep convolutional neural network (CNN)" and the "artificial neural network (ANN)", which have been used accomplish amazing feats such as beating chess grandmasters (Grzybowski Para. 31). In 2022, AI reached a major milestone when an AI chatbot named ChatGPT was released, which many students and teachers now use to answer their questions with human-like responses, though it frequently has problems with accuracy (Grzybowski, Para. 33). All these technological developments combined are used in the creation of SDVs. One example of this is in the study, "Autonomous Traffic Sign Detection for Self-Driving Car System Using Convolutional Neural Network Algorithm" by Zhao Yu and Ting Ye from December 2023, in which the authors use computer vision, deep learning (DL), and CNN to identify traffic signs and estimate accuracy using sample images. This is great for SDVs because it dedicates a whole model to the smaller, singular task of identifying street signs. This makes it easier to find and solve problems because tractable processes within a larger task are separated.

## A Multitude of Decisions

Self-driving cars might seem like they are pre-programmed by a person or group of people to determine the behavior of the vehicle at all times, but that is not the case. It would be impossible to program for all inputs and create outputs for every scenario. Florian Springer describes the way these models work as a virtual "world of probability" (Sprenger, Para. 44) creating a simulation of the world and making predictions based on it. Model behavior is hence up to random chance because of the probabilistic nature of its world model. This matters because of the two-ton metal death machines these programs control. In agreement with Sprenger, digital ethics researcher Dr. Mark Ryan goes in-depth into the legal side of these problematic decisions. Moreover, Ryan brings up the "legal concern that SDVs will be used for malicious, illegal, and fraudulent purposes" (Ryan). This means that companies might need to hire car cybersecurity teams to program safeguards and defense programs to manage the SDVs, which have the potential to kill many people if someone with ill-natured intentions gains control.

On October 2nd, 2024, I interviewed Houjon Liu, a computer science researcher at the Stanford Artificial Intelligence Laboratory. He gave some useful insight into the precautions that these programmers must take when designing and testing AI models with the potential to be used for malicious purposes. More specifically, his research involves DementiaBank, a database storing information regarding dementia patients where they, “have a fairly rigorous IRB process internally to Carnegie Mellon … [as well as] externally through the various providers to deal with this [data security]”. This demonstrates an example of researchers needing to go through an Institutional Review Board (IRB) to ensure that the research they are doing is ethically sound and does not risk harming the patients involved. Similarly, Ryan says "European governments need to effectively integrate the tenets of the GDPR into the automotive industry," meaning that the data protection policies in the European Union could be very useful in integrating with AI to

provide enhanced data security. This could be done by utilizing the already strict rules in the General Data Protection Regulation (GDPR) to protect the new data models being generated by AI (Ryan 1190).

There is a silver lining: the control AI has over the car varies with the intelligence of the car model (defined by the Society of Automotive Engineers (SAE)). Thus, passing a policy to limit the intelligence and control of the car computer will be much easier. Ryan states "In areas where there are mixed drivers (automation and nonautomation), SDVs must have a level 3 option for legal reasons" (level 3 means limited automation). The levels of automation define the responsibility level of the car's human driver while making it clear that the autopilot features must not take over the car's full function. Higher-level autonomous vehicles (AVs) could be used to assist during an emergency in cases in which the human response team is spread thin. According to Clemson University scholar Thomas Shirley's article "Exploring the Potential of Using Privately-Owned, Self-Driving Autonomous Vehicles for Evacuation Assistance," it may be possible to convince a minority of people to contribute their AVs to help people during an emergency such as a flood or fire where people must be evacuated immediately to survive. Overall, using SDVs as emergency response vehicles, changing them to a limited automation mode when in use, using data protection, and improving their security against cyber-based attacks can all be methods to improve the usability and positive contribution of self-driving cars to society.

## Programming Predicament

While the legal and ethical concerns are pressing and valid, the humans who write these models have a connection and commitment to quality that a computer would not. Philosophy professor Jason Millar discusses this conundrum in his article titled "An Ethics Evaluation Tool

for Automating Ethical Decision-Making in Robots and Self-Driving Cars". He appeals to the audience by using examples and convinces the reader that an ethics code is needed for the creation of artificial intelligence models. While Millar does not have any computer science credentials, the philosophical side supports the programmers creating AI models for self-driving cars. In our interview, Houjon Liu spoke about the tools he utilizes in his current research: Adaptive Stress Testing. This method is used to test AI models with sequential decisions where they try “to find likely failure cases that are also bad.” In "Self-driving car dilemmas reveal that moral choices are not universal", Journalist Amy Maxmen discusses a survey called "The Moral Machine", conducted by MIT students in 2018, in which people worldwide were given problems that a machine learning algorithm might have to face. This survey primarily included questions about situations in which a self-driving car must pick between ending the life of one party or another with many variations of the parties on each side. In “The Moral Machine Experiment” by Edmond Awad et al. They include a photo of the interface that the people surveyed used to answer questions, which included a car, some people on the road, and a concrete barricade, with two arrows, one that went to the people and one that went to the barricade (Awad). The results of this survey highlighted the diverse discourse in the world over ethical standards because it demonstrated that different parts of the world have different values and prioritize the lives of some people or pets over others (Maxmen).

More profoundly, Maxmen states that many major car companies have avoided talking about this survey, including Tesla, a car manufacturer with a high focus on self-driving cars. Tesla has reported many recent crashes related to their vehicles’ self-driving features, leading to an investigation by the National Highway Traffic Safety Administration (NHTSA; according to The Verge, a commercial entity (Hawkins)). These crashes were violent and caused many deaths,

even a "17-year-old student" in 2013 (Hawkins). Furthermore, the programmers who worked on the Tesla self-driving project were not questioned when these crashes were brought to court in 2024 for a wrongful death lawsuit, in which a man named Walter Huang was killed by the autopilot driving off the highway at 71 miles per hour (Goldman) according to CNN. In "An Ethical Trajectory Planning Algorithm for Autonomous Vehicles", mechanical engineering doctoral student Maximilian Geisslinger discusses the trajectory an SDV might take when given a scenario and risk values. In the article, he states that putting a threshold on risk will create "on the one hand, better decisions under the assumption of a reliable risk assessment of AVs and, on the other hand, transparency towards all traffic participants concerning the AV's functioning." This highlights how explicitly programming safety into implicit neural networks could resolve ethical decision-making dilemmas and provide an alternative for drivers who want to balance caution and driving efficiency.

Self-driving cars have the potential to shape the future and redefine the way we travel, but the challenges we face in their development should not be overlooked. When computers were first invented by Charles Babbage and Ada Lovelace, they were simple and limited. Modern computing utilizes AI models that can be generalized for a wide variety of situations and respond accordingly without user input. Self-driving vehicles use AI models to accurately predict and respond to different scenarios encountered while driving. There are ways to make these models safe, secure, and usable for everyone. However, they have ethical concerns: these models have to decide whether or not to end people’s lives, an impossible expectation. Philosopher Jason Millar proposed a tool to evaluate these situations based on the philosophy to be used when programming self-driving car AI models. Extensive testing ensures that the models are ethically

sound and do not make bad decisions, but they require more testing: crashes connected to Tesla's new Autopilot feature are a poignant example. The specific goals of the tests have been studied for some time now with a group of students at MIT even surveying it to determine the world's moral priorities. Additionally, the sophisticated sensor technology required to inform SDVs of their surroundings collects a significant amount of data that must be protected by laws like the EU's GDPR. Also, research into SDVs may use an IRB to further ensure proper ethics, allowing an external board of reviewers to give their opinions on the work before it can be undertaken. The best way to go forward with modern SDV algorithms is to keep pushing for better data protection policies and testing methods that ensure that AI-powered driving does not put pedestrians and other vehicles at even more risk.

# Works Cited

Awad, Edmond, et al. “The moral machine experiment.” Nature, vol. 563, no. 7729, 24 Oct. 2018, pp. 59–64, https://doi.org/10.1038/s41586-018-0637-6.

Geisslinger, Maximilian, et al. “An Ethical Trajectory Planning Algorithm for Autonomous Vehicles.” Nature News, Nature Publishing Group, 2 Feb. 2023, www.nature.com/articles/s42256-022-00607-z.

Goldman, David. “Tesla Settles with Apple Engineer’s Family Who Said Autopilot Caused His Fatal Crash | CNN Business.” CNN, Cable News Network, 8 Apr. 2024, www.cnn.com/2024/04/08/tech/tesla-trial-wrongful-death-walter-huang/index.html.

Grzybowski, Andrzej, Katarzyna Pawlikowska–Łagód, and W. Clark Lambert. “A History of Artificial Intelligence.” Clinics in dermatology 42.3 (2024): 221–229. https://doi.org/10.1016/j.clindermatol.2023.12.016.

Hawkins, Andrew J. “Tesla’s Autopilot and Full Self-Driving Linked to Hundreds of Crashes, Dozens of Deaths.” The Verge, The Verge, 26 Apr. 2024, www.theverge.com/2024/4/26/24141361/tesla-autopilot-fsd-nhtsa-investigation-report-crash-death.

Houjon Liu, Personal Interview, 2 Oct. 2024.

Maxmen, Amy. “Self-Driving Car Dilemmas Reveal That Moral Choices Are Not Universal.” Nature, vol. 562, no. 7728, 2018, pp. 469–70, https://doi.org/10.1038/d41586-018-07135-0.

Millar, Jason. “An Ethics Evaluation Tool for Automating Ethical Decision-Making in Robots and Self-Driving Cars.” Applied Artificial Intelligence, vol. 30, no. 8, 2016, pp. 787–809, https://doi.org/10.1080/08839514.2016.1229919.

Muthukrishnan, Nikesh, et al. “Brief History of Artificial Intelligence.” Neuroimaging Clinics of North America, vol. 30, no. 4, 2020, pp. 393–99, https://doi.org/10.1016/j.nic.2020.07.004.

Ryan, Mark. “The Future of Transportation: Ethical, Legal, Social and Economic Impacts of Self-Driving Vehicles in the Year 2025.” Science and Engineering Ethics, vol. 26, no. 3, 2020, pp. 1185–208, https://doi.org/10.1007/s11948-019-00130-2.

Shirley, Thomas, et al. “Exploring the Potential of Using Privately-Owned, Self-Driving Autonomous Vehicles for Evacuation Assistance.” Journal of Advanced Transportation, vol. 2021, 2021, pp. 1–11, https://doi.org/10.1155/2021/2156964.

Sprenger, Florian. “Microdecisions and Autonomy in Self-Driving Cars: Virtual Probabilities.” AI & Society, vol. 37, no. 2, 2022, pp. 619–34, https://doi.org/10.1007/s00146-020-01115-7.

Yu, Zhao, and Ting Ye. “Autonomous Traffic Sign Detection for Self-Driving Car System Using Convolutional Neural Network Algorithm.” Journal of Optics (New Delhi), 2023, https://doi.org/10.1007/s12596-023-01518-x.

John-Nicholas Krinos

ENC 2135 – Fall 2024

Prof. Bronstein

Project 1: Annotated Bibliography Final Draft

September 15th, 2024

Annotated Bibliography: The Life or Death of Self-Driving Car Algorithms

Hamers, Laurel. “Five Challenges for Self-Driving Cars.” Science News, 12 Dec. 2016, www.sciencenews.org/article/five-challenges-self-driving-cars. Accessed 15 Sept. 2024.

Laurel Hamers provides a five-topic summary of the problems surrounding self-driving cars in “Five Challenges for Self-Driving Cars”. The webpage starts with some positive statements, pointing out how they could help “people who can’t operate a vehicle” (Para. 1) and how they were already being tested out on the road when they wrote this article. The next portion talks about sensor problems during weather conditions, specifically “I’ve seen promising results for rain, but snow is a hard one,” (Para. 5) which they quote from a professional in the field. Next, she talks about problems communicating with other people about the car’s actions (Para. 10). The webpage goes deeper into this, talking about how self-driving cars would even communicate with the passengers, such as “How does the car notify a passenger who has been reading or taking a nap that it’s time to take over a task...” (Para. 11) without giving a specific answer on how this could be done. She also throws in the “Moral Machine” (Para. 15) which is a social experiment conducted by MIT online to see what people value more in vehicular collisions. The last parts of the webpage talk about two main things, the security of the massive amount of data being collected by these cars’ sensors and their vulnerability to hackers, with an example given of a Jeep being stopped on the road by a hacker “...wirelessly accessing its

braking and steering via the onboard entertainment system." (Para. 16). This article was produced for informational reasons, published by a nonprofit website called "Science News". In my paper, I will use this article to examine other problems with self-driving car programming, outside of the obvious collision questions. Also, I can use this to propose the importance of the cars' security.

Lin, Patrick. "The Ethical Dilemma of Self-Driving Cars." *Patrick Lin: The Ethical Dilemma of Self-Driving Cars | TED Talk*, Dec. 2015, www.ted.com/talks/patrick_lin_the_ethical_dilemma_of_self_driving_cars?subtitle=en. Accessed 15 Sept. 2024.

Patrick Lin gives the consumer a brief overview of the burning question regarding self-driving car ethics using an animated video titled "The Ethical Dilemma of Self-Driving Cars". The video begins with a cold open, throwing the user straight into the meat of the problem: a scenario in which a driver has a loaded decision to make that forces them to take a side. The next portion of the video goes into a variety of what-if questions that set up the viewer to think about more decisions and put them into the mind of a programmer setting these policies, such as asking the viewer if they would rather have the self-driving car hit a motorcyclist wearing a helmet or one without a helmet (2:02), exploring both options and showing the issues with either one. The video summarizes this example by stating that regardless of the choice, having a decision-making algorithm for this type of machine would be "systematically favoring or discriminating against a certain type of object to crash into." (2:34). The next part of the video is essentially a summary of questions from the field, asking many rhetorical questions such as "What happens if the cars start analyzing and factoring in the passengers of the cars and the particulars of their lives?" (2:58). This video was produced for educational purposes to make the public aware of

these difficult questions and to introduce interdisciplinary discussion of the topic by generalizing the problem for anyone to get a grasp of. I plan to use this source to provide more examples of problems with self-driving cars and to introduce the legal/business problems of self-driving cars.

Maxmen, Amy. “Self-Driving Car Dilemmas Reveal That Moral Choices Are Not Universal.” *Nature*, vol. 562, no. 7728, 2018, pp. 469–70, https://doi.org/10.1038/d41586-018-07135-0.

While people in the U.S. might have certain morals and standards, these standards may differ throughout the world and driving is a great example. A survey was taken of 2.3 million people worldwide about the morals behind programming self-driving cars and the results varied greatly between different regions of the earth. Amy Maxmen analyzes the results in the article "Self-driving car dilemmas reveal that moral choices are not universal" and briefly summarizes the questions given in the survey. Furthermore, it contained "...13 scenarios in which someone's death was inevitable." (Para. 4) which were differentiated by the subjects in the scenarios. The results were made public, but Maxman states that many big tech companies refused to reply to its findings. Next, she starts delving into the results directly, first pointing out the similarities: "...most people spared humans over pets, and groups of people over individuals." (Para. 11). Then, she uses a graphic to illustrate the broad trends of the survey, divided into three groups: Western, Eastern, and Southern. The graphic itself is a "Moral Compass" that has three shapes layered on top of each other for each group and nine points each representing a different value such as "Sparing the young" or "Sparing the fit" to accentuate different responses to the survey questions. She notes that some companies have commented on the survey such as Barbara Wege

from Audi who "...says such studies are valuable." (Para. 19). Finally, she ends the paper with a reflection on the concept of self-driving cars, pointing out their safety problems and how this survey will affect the way we approach the issue in the future. I might use this article in my paper to provide meaningful examples and possible solutions that people may program into their algorithms.

Millar, Jason. “An Ethics Evaluation Tool for Automating Ethical Decision-Making in Robots and Self-Driving Cars.” *Applied Artificial Intelligence*, vol. 30, no. 8, 2016, pp. 787–809, https://doi.org/10.1080/08839514.2016.1229919.

Jason Millar aims to describe an ethical code of conduct for technology in the article “An Ethics Evaluation Tool for Automating Ethical Decision-Making in Robots and Self-Driving Cars”. The article starts with four jarring examples of robots and self-driving cars being faced with difficult decisions that would be hard for even a human to make. The article continues by using these four examples to justify the need for “automating ethical decision-making…” (Para. 8). Next, he lays out the groundwork for his proposed automation, settling on five key rules for its creation ranging from being user-centered to following the existing “…human-robotics interaction (HRI) Code of Ethics.” (Para. 18). Next, he drafts a document using the five rules he has set. He invites the reader to read it by attaching it to the article. Additionally, he includes a short section on governance, or the control of the robots, stating that his ethics evaluation tool will solve both problems because “…governance and design ethics issues surrounding the automation of ethical decision-making will often appear as two sides of the same coin. Finding solutions for one will suggest solutions for the other.” (Para. 40). Lastly, he concludes the article by admitting that it is

all theoretical and he has not done any testing to see if his tool would work if created and applied to real-world scenarios, stating that he needs to validate it and "A validation process will require interdisciplinary collaboration." (Para. 41). The target audience of this article are engineers and designers in the field of robotics who need to consider ethics in their design, shown by the professional tone of the article and the serious proposals it makes. I plan to use this article to provide an example of rules programmers could follow when determining ethics instead of using their moral compasses.

Shirley, Thomas, et al. "Exploring the Potential of Using Privately-Owned, Self-Driving Autonomous Vehicles for Evacuation Assistance." *Journal of Advanced Transportation*, vol. 2021, 2021, pp. 1–11, https://doi.org/10.1155/2021/2156964.

Thomas Shirley wrote the article "Exploring the Potential of Using Privately-Owned, Self-Driving Autonomous Vehicles for Evacuation Assistance", to explore using consumer-owned Autonomous Vehicles (AVs) as emergency response vehicles. He starts the article with some background about the destruction that recent disasters have caused in the U.S., and how the U.S. has struggled to appropriately respond to these threats with the resources they have allocated to evacuation plans (Shirley 1). Next, he breaks down a survey conducted on South Carolina residents into a table showing the number of people surveyed and their opinions on sharing their self-driving cars and their views on self-driving cars in general (Shirley 4). Shirley goes on to explain that certain demographics were sampled more because of their higher likelihood of being able to afford self-driving cars. Finally, he concluded that "…approximately 32% of households were willing to share their AVs to assist with hurricane evacuation." (9). This article is a good

example of the positive applications of a self-driving car and the ethical benefits that could be reaped from programming them to assist people. The author and source are credible because it is from a peer-reviewed journal and the authors all have relevant degrees from reputable colleges in the U.S. The intended audience of this paper is scholars researching self-driving cars and their applications, as shown by the extensive detail on the statistical side of the paper and the great amount of data provided. I imagine using this article in the middle of my essay to give a counterexample to the deadly threat of self-driving cars.

Sprenger, Florian. "Microdecisions and Autonomy in Self-Driving Cars: Virtual Probabilities." AI & Society, vol. 37, no. 2, 2022, pp. 619–34, https://doi.org/10.1007/s00146-020-01115-7.

Florian Sprenger pursues a redefinition of the concept of self-driving car algorithms, and their actual autonomy. To begin with, he talks about the advancements made in autonomous technology and how the focus of the world has turned to the "...Advanced Driver Assistance Systems (ADAS) that are installed in all new cars..." (Para 1) and their importance. The point he emphasizes is that computers are evolving, they are now taking in data from their surroundings using sensors and are interacting with their environment with robotics and machines, which is not something that can be predicted in the way that a computer program can predict all possible inputs. However, Sprenger does not believe that these computers are at a "human-like autonomy" (Para 3) now and are just using "microdecisions" (Para. 3) to account for their lack of autonomy. His explanation for how this works is that they are using probability models to make these micro-decisions and that "they are no representations of the world" (Para 5) speaking to the

inaccuracies of these models. Continuing, he uses a graphic made by the Society of Automotive Engineers (SAE) to categorize the capacity of a self-driving car into five levels, ranging from zero or no automation to five or full automation. He states that the sensor hardware to accomplish the higher levels of automation is already installed into most car models and is just waiting for "new software" to allow it to happen (Para. 15). Next, he explains how the cars don't see the real world, only their virtual "world of probability" (Para. 44) and details the possible effects of this, "the system is capable of creating new internal worlds because it has no access to the outside world" (Para.48). Lastly, he reinforces what he said in the beginning about the vocabulary we use in discourse about self-driving cars, stating that the word micro-decision is better to use than algorithm and reiterates the definition and composition of these micro-decisions. This paper was written for a scholarly audience, filled with high-level vocabulary and industry references like the SAE. I can use this in my paper to better describe the idea of a self-driving car when I introduce the concept at the beginning and connect it to the level of self-driving that we have accomplished in the present day.