

## **A Code of Ethics for Self-Driving Vehicles**

**Magdalene Gleadow**

### **Abstract**

As self-driving technology advances, relevant ethical frameworks must be developed to guide the engineering profession. This paper provides a deeper understanding of the ethical issues associated with self-driving vehicles. It reviews the relevant literature which makes it clear that there are several key ethical issues which must inform the development of a code of ethics specific to this field. These can be grouped into the four key principles: responsibility, safety, transparency and sustainability. The proposed code of ethics is then applied to a specific case study to evaluate it. The case study introduced is that of self-driving cars as rental cars with specific reference to the New Zealand context. This case study provides a complex real-world situation for the proposed code to be evaluated with. Applying the code of ethics to this case study indicates that the proposed code of ethics offers a relevant ethical framework to guide the engineering profession in self-driving vehicles.

*Keywords:* Self-driving vehicles; Rentals; Ethics.

### **1. Introduction**

As more resource is devoted to the design and development of self-driving vehicles, further applications are understood and communicated, and the demand for these vehicles increases, the requirement for a clear code of ethics becomes critical. In turn, this code must be relevant to, and adhered to, by designers, engineers and all others involved in developing, designing and delivering self-driving vehicles.

Self-driving vehicles are vehicles such as cars, trucks or buses which use vehicle automation and AI. This means that such a vehicle can sense its environment and make decisions and move autonomously (with little to no human input) [1]. Autonomy in vehicles is usually categorised in six levels according to the Society of Automotive Engineers (SAE); these range from Level 0 (fully manual) to Level 5 (fully autonomous). This system is used by several countries, such as the US [2].

Every day, humans drive standard vehicles and modes of transportation not only according to the vehicle and road rules but also with a sense of ethics and responsibility [3]. Whether a vehicle is driven by a human or machine, it will encounter situations in which the choices made may have significant consequences for other road users. Sometimes those consequences will include potential loss of human life. For an autonomous vehicle to be able to operate responsibly, it must deliver an improvement upon the human decision-making process and will need to operate in a way which exhibits a sense of ethics.

#### **1.1. Objective**

The objective of this paper is to discover and discuss the ethical issues that the use of self-driving vehicles raises, and then present and apply a proposed code of ethics. This code is informed by a review of relevant literature to uncover the key ethical issues associated with self-driving vehicles, and to identify best practice in this field. A code of ethics is

outlined and evaluated through its application to a specific case study, namely self-driving vehicles as rental cars.

## 2. Literature review

This section reviews relevant literature to analyse and discuss the main ethical issues that self-driving vehicles pose and to identify best practice when it comes to addressing these issues.

### 2.1. Ethics issues

Much of the literature raises the thought experiment the “Trolley Problem”, applying it to self-driving situations [4, 5, 6, 7, 8]. In this problem, a trolley on a rail is approaching a switch. If it continues its track, it will overrun a group of people. A human operator is given the option to pull the switch to move the trolley to another track, where it will instead overrun one person. The operator has a moral dilemma; do they let five people die or choose to sacrifice the one - in which case they are directly responsible [4]. The idea of self-driving vehicles prioritising situations which will result in less deaths than the alternatives, and the discussion around letting people die (passive) vs. killing someone (active) are heavily discussed [5]. Some of the literature does not accept that this is a relevant real-world engineering problem and argues that the “Trolley Problem” is about people and not vehicles. It asks a human to decide the answer, whereas we cannot program a solution for specific situations and instead the vehicle will learn from training [6].

Discussions of the “Trolley Problem” are often tied into the common issue of a self-driving vehicle's priorities. How should a self-driving vehicle prioritise situations? What should it value more: human life, the law or the vehicle itself [3]? In collision situations, should it choose to hit the cyclist over crossing a double yellow line? Should it prioritise the human occupants' safety over that of pedestrians and other drivers [7]? These are all questions posed by the literature, which includes discussions of the specific situations and possible outcomes. Some writers consider whether self-driving cars should be allowed to make explicit moral judgements. Are artificial agents without free will truly able to exhibit moral behaviour [9]? Instead of letting the vehicle decide, there may be situations in which the system should hand-off the decision-making to the human occupant of the vehicle [7, 9]. The literature proposes that the ultimate answer for self-driving vehicles making moral decisions depends upon whether society sees these machines as advanced cars or as robots [7].

If we allow the vehicle to make moral decisions, what will be required and how will humans trust these entities to make such decisions? Gaining this trust will require transparency of the vehicles' programmed priorities and the confidence that decisions made in critical moments are reasonable, ethical and acceptable to society [7].

If a vehicle makes a decision which results in any legal complications, these decisions need to be justifiable and reproducible [9]. While a human can be held accountable for their judgements and decisions, a machine cannot. Instead, the responsibility will be on the designers, developers, deployers and/or users. As stated in *The Moral Responsibility for Computing Artifacts: The Rules*, “The people who design, develop, or deploy a computing artifact are morally responsible for that artifact, and for the foreseeable effects of that artifact. This responsibility is shared with other people who design, develop, deploy or knowingly use the artifact as part of a sociotechnical system.” [8].

A key issue of importance discussed in the literature concerns the ethical obligation for self-driving vehicles to be safe. A statistic raised is that 94% of crashes are caused by

human error [6, 8]. Therefore, removing human error should decrease rate of crashes. Human drivers may be forgiven for making an instinctive but nonetheless bad split-second decision, such as swerving into incoming traffic instead of turning the other way into a field. However, the engineers of a self-driving vehicle do not have that luxury. They have the time to get it right and therefore bear more responsibility for negative outcomes [5]. This poses the question, should self-driving cars have a higher moral responsibility? Or does that higher responsibility lie with the engineers?

In the case of these vehicles, there are several different types of safety which must be taken into consideration. These include the technical safety elements such as collision avoidance or mitigation, as well as the security safety factors which include cybersecurity, privacy and data protection.

The vehicle intercommunications required for some of the most beneficial features of self-driving cars also pose a significant security risk. Vehicles should be designed to mitigate these risks through something such as an intrusion detection system [10].

Once again, we are brought back to a reflection and decision on what self-driving vehicles need to prioritise. There may be situations in which a choice must be made between several options which lead to different negative or even lethal outcomes for one or more of the parties involved [4]. As outlined in the first principle of the Engineering NZ code of ethics, engineers must “*take reasonable steps to safeguard health and safety*” [11]. One component of this is understanding the risks to health and safety along with the possible and likely consequences.

In the case of physical safety, the consequences could be death or serious injury for the user or other parties involved. In the case of data security, if location information combined with personal information is leaked, a person’s details such as wealth status, profession, sexual association, and religion can be deduced. If present and/or historical travel patterns are leaked, then the user could be susceptible to physical harm or stalking [12]. As outlined in the third principle of the Engineering NZ code of ethics, engineers must “*Maintain confidentiality*” [11]. These privacy and trust issues are a key factor in the design and development of self-driving vehicles.

Another important issue is sustainability. With any vehicle design, it is imperative that the vehicle's emissions and impact on the planet are considered. Several pieces of the literature examine this issue. One states that there is a link between broad sustainability themes and self-driving vehicles [7]. Self-driving technology can monitor traffic flow, road markings, lights and signs as well as other road users. All of this means that the vehicle can make more informed and preferred decisions. One piece of literature discusses the particulars of whether current self-driving cars have lower emissions than those generated by a human-driven car [13]. Overall, much of the literature supports the development and use of self-driving vehicles as a key component of the mission to achieve more sustainable forms of transport for our societies and our planet.

### **3. Code of Ethics**

The engineering profession has developed its own code of ethics, and the general principles laid out in this framework must be considered in this field as in others.

A review of the literature makes it clear that there are several key issues which must inform the development of a code of ethics specific to this field – the development and delivery of self-driving vehicles. These can be grouped into the four key principles: responsibility, safety, transparency and sustainability.

These principles must be considered before, during and after the production of a self-driving vehicle.

### **3.1. Responsibility**

The first principle of the proposed code of ethics concerns responsibility. This can be seen as a key principle in two areas:

- firstly in the making of decisions during the design phase, including key moral decisions; and
- secondly, in ensuring that the design, and in particular key safety features, is delivered, audited and that the end product is as specified.

There will be times when decisions must be made regarding a self-driving vehicle, and these decisions will have consequences. These decisions and consequences may result in legal proceedings, and certain steps will be followed to determine liability and responsibility. A self-driving vehicle cannot be held 'morally' responsible or be held accountable in the same way a human can.

Any decisions that the self-driving vehicle makes should be justifiable and reproducible. There must also be a clear understanding for all involved as to the possible consequences and to what extent each person is liable.

The process of designing a self-driving vehicle is complex and highly technical. All steps in the design process must be audited. In the case of any incident, there should be reasonable evidence to find where the issue arose from. This is important not only for responsibility but also for improvement of the system. There must be a clear hierarchy of human responsibility for the system and any fault.

### **3.2. Safety**

The second principle of the proposed code of ethics concerns safety. A good self-driving vehicle needs to be safe, accurate, reliable and secure. Those designing and developing a self-driving vehicle must take reasonable steps for the vehicle to drive safely and responsibly. The self-driving car should be approached with a safety by design process.

The self-driving vehicle's performance should be evaluated to validate its accuracy and safety. This includes its collision avoidance and mitigation features.

The self-driving vehicle should be reliable; this means it should do what is expected and intended and should be consistent.

The self-driving vehicle handles private information and must maintain the data privacy and integrity to be secure.

Design considerations such as obvious kill-switches are required.

### **3.3. Transparency**

The third principle of the proposed code of ethics concerns transparency. The self-driving vehicle will have set priorities and will be trained to deal with situations in a specific way. The programmed priorities should be transparent to the user and society.

It is important that all possible consequences, such as death, serious injury and breach of privacy, are understood by all parties involved. Human-informed consent is necessary.

The self-driving vehicle should communicate in real-time to the driver the decision and action it is taking.

### **3.4. Sustainability**

The fourth principle of the proposed code of ethics concerns sustainability. Self-driving vehicles should be designed with sustainability as a key consideration. Emissions, environmental impact and how the vehicle drives efficiently must be considered in this process.

## **4. Case Study Discussion**

This section of the paper will introduce a specific context case study to which we will apply the proposed code of ethics.

The case study introduced is that of self-driving cars as rental cars with specific reference to the New Zealand context. Due to the additional complexity, an examination of this case study will allow us to test the proposed code of ethics and evaluate it.

In this case study, we have an additional party involved and therefore a more complex issue as to how responsibility should be allocated. The owner of the vehicle and driver will not be the same entity. Furthermore, the owner is a corporation rather than an individual. Many of the decisions they make will have significant impact on all aspects of this code of ethics.

While all vehicle purchase decisions include a consideration of price, in this case study we also have a corporation which is charged with delivering a profit to its shareholders. This may incentivise companies to cut corners and seek out cheaper options which are not appropriate for this specific market, and which do not comply with this code of ethics.

Secondly in the case of rental cars the drivers will often have quite different attributes from the average driver within New Zealand. This is due to tourists and foreigners being a larger part of the rental car market. Drivers of rental cars may:

- not know the New Zealand road rules;
- be unfamiliar with driving on the left-hand side of the road;
- have a language barrier; and
- have varying levels of driving competency.

### **4.1. Responsibility**

Responsibility for the design phase will continue to sit with the engineers, who need to ensure that there are self-driving cars which are appropriate for this market. For example, able to communicate in multiple languages etc. While responsibility in the design phase of the self-driving vehicle lies with the engineer/designer, there is also a responsibility for the owner (rental car company) to purchase a self-driving car which is appropriate for this specific market.

Regarding the second area of responsibility, there may need to be specific regulations targeting the rental car market, for the cars to be audited against. This responsibility will be shared with lawmakers in New Zealand. An additional party will be the insurance companies. Their requirements will inform who can rent and drive a rental car.

The rental car company as owner will be responsible for maintenance and ongoing audits concerning the vehicle.

The rental car agreement will need to make it clear where responsibility for certain decisions and responses may lie. For example, if the vehicle makes a questionable decision then the driver may be responsible for flagging this up to the rental car company.

#### **4.2. Safety**

The rental car company must ensure that they purchase safe, accurate, reliable and secure self-driving cars. Self-driving vehicles provide an opportunity to mitigate some of the existing safety risks associated with rental cars and foreign drivers in New Zealand. As discussed earlier, self-driving vehicles lower the rate of collisions as they remove the human error. This is particularly relevant when drivers may have little understanding of local road conditions. Given these safety benefits for this market an additional ethical issue is whether foreign drivers should be able to hire a non-self-driving rental vehicle.

Safety considerations around the data privacy in the case of rental vehicles include additional factors such as sharing of information between the owner (rental car company) and the driver and ensuring that private information is wiped between different renters of the same vehicle.

#### **4.3. Transparency**

Some of the information provided to the owner of the vehicle must be passed on to the driver, such as the programmed priorities. The moral judgements made by the rental car company and New Zealand society (regulators) may be different from those the individual might make or expect in their own culture.

In the case of rental cars, self-driving vehicles provide extra benefits in terms of transparency, such as communicating with the driver in different languages, explaining the road rules and more.

#### **4.4. Sustainability**

It is important that the rental car company is incentivised and even required to purchase vehicles which comply with these ethical guidelines rather than simply prioritising profit margins. Historically vehicles which impact negatively on the environment have often been cheaper to purchase.

### **5. Conclusion and Recommendations**

The development and design of self-driving vehicles is an ever-growing industry. It is imperative as this technology continues to develop, that the ethical and moral decisions involved are carefully considered. The requirement for a code of ethics to inform this industry is critical. It is highly recommended that time is put into ensuring that these codes are relevant and adhered to. Self-driving vehicles offer enormous benefits to our society. For them to operate to the fullest extent, society must be able to trust these machines and have confidence in the code of ethics that they adhere to. This paper discovered and discussed the key ethical issues that the use of self-driving vehicles raises. It has presented a four-principle code of ethics covering responsibility, safety, transparency and sustainability. It has then applied this code of ethics to a relevant case study - the use of self-driving vehicles as rental cars. This has indicated that the proposed code of ethics is helpful and appropriate in even this highly complex aspect of the use of the technology.

### **References**

[1] What is an Autonomous Car? Synopsis.com.  
<https://www.synopsys.com/automotive/what-is-autonomous-car.html> (accessed Apr. 30, 2022)

- [2] SAE Levels of Driving Automation. SAE J3016. SAE International. May 2021. [Online]. Available: <https://www.sae.org/blog/sae-j3016-update>
- [3] S. Pan, S. M. Thornton and J. C. Gerdes, "Prescriptive and proscriptive moral regulation for autonomous vehicles in approach and avoidance," *2016 IEEE International Symposium on Ethics in Engineering, Science and Technology (ETHICS)*, 2016, pp. 1-6, doi: 10.1109/ETHICS.2016.7560049.
- [4] P. Wintersberger, A. -K. Prison, A. Riener and S. Hasirlioglu, "The experience of ethics: Evaluation of self harm risks in automated vehicles," *2017 IEEE Intelligent Vehicles Symposium (IV)*, 2017, pp. 385-391, doi: 10.1109/IVS.2017.7995749.
- [5] P. Lin, "The ethics of autonomous cars," *The Atlantic*, Oct. 8, 2013, <https://www.theatlantic.com/technology/archive/2013/10/the-ethics-of-autonomous-cars/280360/>, accessed Apr. 23, 2022.
- [6] T. Holstein, G. Dodig-Crnkovic and P. Pelliccione, "Real-world Ethics for Self-Driving Cars," *2020 IEEE/ACM 42nd International Conference on Software Engineering: Companion Proceedings (ICSE-Companion)*, 2020, pp. 328-329
- [7] J. C. Gerdes and S. M. Thornton, "Implementable Ethics for Autonomous Vehicles," in *Autonomous Driving: Technical, Legal and Social Aspects*, W. Wachenfeld and M. Maurer, Eds. Springer (in press), 2015.
- [8] J. Borenstein, J. Herkert and K. Miller, "Self-Driving Cars: Ethical Responsibilities of Design Engineers," in *IEEE Technology and Society Magazine*, vol. 36, no. 2, pp. 67-75, June 2017, doi: 10.1109/MTS.2017.2696600.
- [9] L. D. Riek, D. Howard, "A Code of Ethics for the Human-Robot Interaction Profession," 2014. [Online]. Available: [https://papers.ssrn.com/sol3/s.cfm?abstract\\_id=2757805#references-widget](https://papers.ssrn.com/sol3/s.cfm?abstract_id=2757805#references-widget)
- [10] J. Straub et al., "CyberSecurity considerations for an interconnected self-driving car system of systems," *2017 12th System of Systems Engineering Conference (SoSE)*, 2017, pp. 1-6, doi: 10.1109/SYSOSE.2017.7994973.
- [11] Code of Ethical Conduct. Engineering New Zealand. [Online]. Available: <https://www.engineeringnz.org/engineer-tools/ethics-rules-standards/code-ethical-conduct/>
- [12] T. -H. Nguyen, T. G. Vu, H. -L. Tran and K. -S. Wong, "Emerging Privacy and Trust Issues for Autonomous Vehicle Systems," *2022 International Conference on Information Networking (ICOIN)*, 2022, pp. 52-57, doi: 10.1109/ICOIN53446.2022.9687196.
- [13] M. Massar, I. Reza, S.M. Rahman, S.M.H Abdullah, A. Jamal, F.S. Al-Ismael, "Impacts of Autonomous Vehicles on Greenhouse Gas Emissions—Positive or Negative?," *Int. J. Environ. Res. Public Health* 2021, pp.19, doi: 18.5567. <https://doi.org/10.3390/ijerph18115567>.