

Government's Response to Ethical Dilemmas in Autonomous Vehicle Accidents: An Australian Policy Evaluation

Amir Rafiee^{1,3}, Hugh Breakey², Yong Wu^{3,4} and Abdul Sattar^{1,4}

¹*School of Information and Communication Technology, Griffith University, Brisbane, Australia*

²*Griffith University's Institute for Ethics, Governance and Law, Brisbane, Australia*

³*Griffith Business School, Griffith University, Gold Coast, Australia*

⁴*Institute for Integrated and Intelligent Systems, Griffith University, Brisbane, Australia*

{*amir.rafiee, h.breakey, yong.wu, a.sattar*}@griffith.edu.au

Keywords: Autonomous Vehicle, AI Policy, AI Ethics, Automated Driving System, AI Regulation, Safety Assurance System, Self-Driving Cars, Ethical Dilemmas.

Abstract: As Autonomous Vehicles (AVs) rapidly progress and become widely deployed, governments worldwide grapple with addressing the ethical challenges associated with AVs in dilemma situations that result in loss of human life. They are tackling these issues through the formulation of policies and guidelines, the establishment of dedicated research centres exploring the ethical implications of AVs, and seeking public opinions on how self-driving cars should handle such moral dilemmas. In this paper, we will evaluate the Australian government's strategies for addressing the ethical issues related to AV accidents. We will critique the Decision Regulation Impact Statement (DRIS) released by the National Transport Commission (NTC) in 2018, which assessed the safety assurance options for Automated Driving Systems (ADSs). We will critically examine the NTC's decision to exclude ethical considerations for AVs from the DRIS and the Automated Driving System Entity's (ADSE) Statement of Compliance. We will contend that safety and ethics are intrinsically linked. Furthermore, we argue that relying solely on the safety criteria may prove inadequate when addressing ethical dilemmas. Consequently, we advocate for the inclusion of AV ethical considerations, especially in complex ethical dilemmas, in future dialogues, even if a clear consensus on ethical decision-making by ADSs remains elusive. In conclusion, we will propose recommendations for the Australian government to enhance the development, deployment, and acceptance of AV technology.


1 INTRODUCTION


The anticipated impact of Autonomous Vehicles (AVs) on the road transport system, including improvements in safety, productivity, environmental sustainability, mobility, and accessibility, has led to efforts by countries worldwide to develop regulatory frameworks that enable the legal and safe operation of this technology. Additionally, governments are motivated to position themselves as leaders in promoting emerging technologies and industries that can create job opportunities and stimulate export revenues (NTC, 2022).


However, the introduction and use of AVs also in-


troduce new risks, which are particularly concerning due to the emerging and evolving nature of this technology. Software malfunctions have been identified as a potential risk associated with AVs (Noy et al., 2018). Safety engineers anticipate that systemic technical failures or inadequate maintenance and servicing of the Automated Driving System (ADS) could become serious hazards, much like human errors (Kira, 2017). Furthermore, governments worldwide grapple with the ethical challenges of AVs in morally complex situations, such as ethical dilemmas. These arise when an AV faces a decision that could potentially harm individuals, such as choosing between the safety of its passengers or the safety of pedestrians in a collision (FCAI, 2023).

Governments are tackling these issues through the formulation of policies and guidelines, the establishment of dedicated research centres exploring the ethical implications of AVs, and seeking public opinions

^a <https://orcid.org/0000-0002-8149-4780>

^b <https://orcid.org/0000-0002-5170-2091>

^c <https://orcid.org/0000-0003-2206-8594>

^d <https://orcid.org/0000-0002-2567-2052>

on how self-driving cars should handle ethical dilemmas in accidents that result in loss of human life. These efforts aim to ensure the safe, ethical, and socially responsible development, deployment, and regulation of AVs, all while considering the cultural, legal, and societal variations across different countries and regions (Krueger et al., 2016).

In this paper we will explore the regulatory regime governing AVs in Australia and evaluate the Australian government's response to address ethical challenges linked to AV accidents, especially in dilemma situations. Resolving ethical dilemmas is crucial for ensuring safety, fostering public trust, complying with legal frameworks, maintaining moral responsibility, shaping public perception, and preventing unintended consequences. Addressing these challenges is therefore, fundamental to the ethical and responsible development and deployment of AV technology (Rafiee et al., 2023). The second section of our analysis begins by looking at the safety assessment criteria mentioned in the National Transport Commission's (NTC's) Decision Regulation Impact Statement (DRIS), which was published in 2018. We will critically examine the rationale behind their decision to exclude ethical considerations from the DRIS. The NTC argued that a separate ethical considerations criterion is unnecessary, as such considerations are already captured by the proposed safety criteria. Their concern was that introducing a distinct ethical criterion might result in redundancy and potential conflicts with the established safety criteria. The NTC also referenced a solution proposed by Swedish academics, suggesting that if an ADS can assess its operational capability to handle unexpected situations and adapt its behavior, it could effectively address ethical dilemmas (NTC, 2018; Johansson and Nilsson, 2016). We contend that effective management of safety risks necessitates the inclusion of ethical considerations in future discussions, and we present arguments supporting this stance. In conclusion, we will make some recommendations for the Australian government and policy makers to better support the development, deployment, and acceptance of AV technology.

2 SAFETY ASSURANCE FOR AUTOMATED DRIVING SYSTEM

In 2018 the NTC published the report, "Safety Assurance for Automated Driving Systems: Decision Regulation Impact Statement". The aim of the DRIS was to evaluate potential approaches for ensuring the

safety of ADS. This involved an analysis of safety concerns related to AVs and an examination of possible government actions in response to these concerns. The DRIS used a multi-criteria analysis to assess options for the safety assurance of ADSs. This approach enabled the assessment of both quantitative and qualitative information against criteria linked to expected impacts, encompassing costs and benefits associated with various policy options. Given the uncertain nature of future of AVs and the lack of empirical data, a fully quantitative cost-benefit assessment was deemed impractical. The analysis focuses on road safety, AV uptake, regulatory costs, and technology flexibility, assigning weight based on materiality tests that prioritise significant road safety and economic benefits resulting from increased adoption. The assessment spanned a 10-year timeframe, employing a comparative scale to rate options against a baseline. The validity of outcomes was tested against different deployment scenarios, ensuring a robust framework for evaluating policy impacts on automated vehicle safety assurance (NTC, 2018).

The DRIS outlined four options to address the safety assurance of AVs. The baseline option (option 1) did not develop a novel safety assurance system. Option 2 enhanced the existing regulatory framework by introducing a safety assurance system, requiring Automated Driving System Entities (ADSEs) to self-certify. This option was updated based on feedback received during the consultation process. Option 3 suggested a nationally administered safety assurance system, applied at both the first supply and in-service stages. It includes specific offenses and compliance measures against non-compliant ADSEs. Option 4 incorporated all elements of Option 3 and introduced a "primary safety duty", imposing a general duty on ADSEs to ensure safety. Stakeholder input has informed updates to these options, ensuring a comprehensive consideration of safety assurance measures for automated vehicles. After considering various options through consultations, option 2 was recommended.

2.1 Mandatory Self-Assessment

There are eleven safety criteria that the ADSEs must self-certify against to demonstrate their processes for managing safety risks before their ADS can be supplied in the Australian market. In examining regulatory options for AVs, the assessment centered on their adaptability to uncertainties. The criteria considered encompassed the feasibility of implementation by 2020, in accordance with industry forecasts and governmental goals. The evaluation also factored

in the ability to transition in tandem with evolving international regulations. These criteria were developed with the aim of balancing safety and innovation. Additionally, the assessment acknowledged the dynamic development of ADS technology, necessitating regulatory responsiveness to emerging safety risks (NTC, 2018).

The safety criteria have been integrated into the existing framework for the initial supply of vehicles under the Road Vehicle Standards Act (RVSA) 2018. The administration of this framework falls under the Department of Infrastructure, Transport, Regional Development, and Communications (DITRDC). Furthermore, DITRDC is currently in the process of incorporating these safety criteria into the Australian Design Rules (ADR 90/01). Consequently, there have been no further amendments to the safety criteria for initial supply as outlined in the NTC's DRIS, affirming its ongoing relevance. Full descriptions of the safety criteria and obligations, are included in (Appendix .1).

The safety criteria are considered to be principle-based since they are formulated as overarching principles rather than specific, prescriptive rules. Principle-based regulations provide a framework of broad principles and goals that guide behavior and decision-making, allowing for innovation and adaptation to evolving technology. These principles are intended to guide the development, deployment, and regulation of AVs in a manner that prioritises safety without being overly restrictive. They accommodate technological advancements and allow for a more agile response to emerging safety concerns and advancements in technology without the need for frequent regulatory revisions.

In this section, we provide a summary of the four key principle-based safety criteria set by the government for the Statement of Compliance (SOC). The SOC is a mechanism for transparency and accountability, demonstrating that the ADSE takes responsibility for the safety of its ADS and that it aligns with the principles and requirements set forth in the regulatory framework. We focus on these principles as they are intended to ensure the safe operation of AVs and are especially important to the decision-making aspects of AVs in ethical dilemmas. The NTC suggests that the safety criteria can serve as a framework for addressing safety dilemmas with ethical implications. Consequently, they deem a separate criterion specifically for ethical considerations as unnecessary. However, they also acknowledged that these safety criteria may not assist the ADS in making decisions involving human lives and emphasised that such decisions cannot be programmed into AVs regardless (NTC, 2018).

1. **Safe System Design and Validation Processes.** Applicants must provide a comprehensive justification for their chosen design, validation, and verification procedures, emphasising the system's safety throughout its entire operational life. The applicant must also address safety-critical concerns such as maintenance, repairs, system failures, and the end-of-life phase of the ADS when support is no longer available. Additionally, if the ADS is intended as an aftermarket device, it should clearly specify compatibility with different vehicle types in its design.
2. **Operational Design Domain (ODD).** Applicants must identify the ODD of the system and demonstrate its ability to operate safely within this defined domain. It should also show that the ADS cannot operate outside of its defined ODD and has the capability to transition to a minimal risk condition if it finds itself outside this domain. For example, if the ADS cannot safely operate at night or in certain weather conditions, then these should be outside the scope of its defined ODD, and the ADS should be incapable of operating in such conditions (NTC, 2018).
3. **Minimal Risk Condition.** The ADS may be unable to operate safely where there are system faults, including as a result of a crash, where there is a deterioration of vehicle hardware or physical tampering, or where the ADS is outside its ODD. Applicants are obligated to provide evidence of the ADS's capability to recognise such situations where safe operation is not feasible, along with the measures it will employ to mitigate risks.
4. **On-road Behavioural Competency.** Applicants are required to demonstrate how the ADS will respond effectively to various situations, to ensure safe operation. This involves actions like disengaging the system or safely stopping the vehicle. The documentation should also outline how the system will interact safely with other road users, covering areas such as detecting objects, avoiding accidents, handling unusual or unexpected situations within its ODD, and interacting with different road users, including those who may be vulnerable.

2.2 Exclusion of Ethical Considerations

In the DRIS, the NTC explained that while they recognise that ADSs might encounter safety dilemmas with ethical implications, they do not consider it necessary to include a separate ethical considerations criterion as they are already largely captured by

the safety criteria. They argued that introducing a separate ethical considerations criterion might lead to substantial redundancy and possible conflicts with the existing safety criteria (NTC, 2018). The NTC also highlighted that there might be a potential solution for addressing safety dilemmas with ethical implications in ADSs, referencing a conference paper by two Swedish academics (Johansson and Nilsson, 2016). This solution suggests that if an ADS can evaluate its operational capability to handle unexpected situations and adjust its behaviour accordingly, it might help resolve ethical dilemmas. Consequently, they decided to exclude ethical considerations from their discussions.

3 DISCUSSION

In this section we will critically examine the NTC's decision to exclude ethical considerations for AVs from the DRIS and the ADSE's SOC. We will assess each argument and offer our perspective on why these rationales may not be adequate to justify the exclusion of ethical considerations from future discussions. We will contend that safety and ethics are inherently connected, with ethics playing a vital role in safety by establishing acceptable levels of harm. Furthermore, we argue that relying solely on the safety criteria may prove inadequate when addressing ethical dilemmas, especially considering that the capabilities of forthcoming AVs at levels 4 and 5 of automation have been overly exaggerated. These are vehicles where no driver is ever needed, or there might be an option for human override but not a requirement (Shadrin and Ivanova, 2019). Consequently, we advocate for the inclusion of AV ethical considerations as part of future discussions, particularly when dealing with complex ethical dilemmas. It is imperative to establish this as a mandatory criterion, even if a clear consensus on ethical decision-making by ADSs remains elusive.

3.1 Argument-1 "Ethical Considerations Captured by Safety Criteria"

The initial argument revolved around the NTC's perspective that ethical concerns were mostly addressed by the safety criteria. These criteria require ADSEs to self-certify, demonstrating how they manage safety risks before introducing their ADS to the Australian market. They argued that introducing a separate ethical considerations criterion might lead to substantial

redundancy and possible conflicts with the existing safety criteria (NTC, 2018).

Firstly, we define moral dilemmas in the context of AVs. Subsequently, we present a hypothetical scenario illustrating such dilemmas. While this example may seem improbable, it remains conceivable and thus merits examination. We will then assess the adequacy of the four safety criteria emphasised by the NTC in the DRIS for addressing such situations.

Subsequently, we shall assess whether these criteria align with specific objectives aimed at ensuring the ethical acceptability and fairness of AVs (Rafiee et al., 2023). This evaluation holds significance, especially in achieving widespread AV adoption, which is necessary in unlocking the potential benefits associated with fully autonomous vehicles on our roads (Maurer et al., 2016).

3.1.1 Ethical Dilemmas Defined

An ethical dilemma refers to a situation in which the ADS faces a complex moral decision that may result in the loss of human life, and there is uncertainty or disagreement about the right course of action. These dilemmas arise from the need for the self-driving car to make split-second decisions in real-world scenarios where there may be conflicting ethical principles or values. For instance, should it prioritise protecting its passengers even if it means causing harm to pedestrians? There might be other situations where the ADS faces conflicting priorities. Exceeding the speed limit for example to avoid a dangerous situation could be viewed as an ethical decision that conflicts with legal compliance (Nyholm and Smids, 2016).

As the development of AVs progresses, addressing these ethical dilemmas becomes a critical aspect of ensuring the responsible and ethical deployment of AV technology. However, reaching a consensus on the best ethical framework is challenging, as societal values and individual preferences may vary. The dilemma scenarios would require manufacturers to incorporate regulatory rules for the behaviour of AVs during programming. The ethics programming would provide instructions on how to respond to situations when an accident appears imminent (Wu, 2020). This is not simply pre-setting the conditions, but these vehicles must show behavioural patterns in situations of necessity, where protecting one interest unavoidably inflicts harm on another (Coca-Vila, 2018).

3.1.2 The Safety Criteria Conundrum

In a highway scenario where an AV is closely followed by a motorcyclist, an ethical dilemma arises when a child unexpectedly appears between parked

cars. Faced with only two options, the AV can either engage the brakes to protect the child, resulting in a deadly collision with the motorcyclist, or maintaining its course, causing harm to the pedestrian to spare the motorcyclist. This crash scenario demonstrates a situation where the AV must make instant decisions without the possibility of manual takeover, common in levels 4 or 5 of driving automation. Achieving a minimal risk condition, where all parties avoid serious harm, is challenging in such circumstances, emphasising the complexity of ethical decisions in AV technology.

To enable an ADS to effectively respond to unexpected situations, the ADSE must demonstrate its ability to anticipate and plan for a wide range of potential crash scenarios. However, it is essential to acknowledge that, as evidenced by incidents such as the Tesla Model S and Uber Technologies crashes in 2016 and 2018 respectively, anticipating every conceivable crash scenario is nearly impossible (Yadron and Tynan, 2016). Technological progress occurs incrementally, and it would be unrealistic to expect flawless AV deployment. Drawing parallels with the development of modern commercial airliners, which took decades to achieve nearly zero airline crashes through the gradual improvement of autopilot systems and computer integration (Dewilde, 2022), emphasises the time and evolution required for advancing AV technology.

The ADS faces additional challenges compared to modern commercial airliners due to other factors, such as complex roadways, variable speed limits, changing environmental conditions, and the presence of manually operated vehicles. These factors, combined with the ADS's ability to independently identify and manage risks, introduce further complexity. Such variability might result in ADSs reacting differently to identical dilemma situations, creating unpredictability for both other ADSs and fellow road users.

We have argued that presuming future AVs at levels 4 and 5 of automation are immune to crashes and moral decision-making scenarios is improbable. Even if the ADSE adheres to the NTC's safety criteria, it is crucial to assess their ADS against specific objectives proposed by some authors to ensure ethical acceptability and fairness (Rafiee et al., 2023). Firstly, ensuring transparency regarding programmed priorities and assigning blame in case of accidents is crucial. This transparency not only informs potential AV customers about how the AV will handle moral dilemmas but also allows policymakers to anticipate and address legal implications. Secondly, the ADSE must minimise public outrage and discrimination by demonstrating that their ADS treats all individuals fairly, re-

gardless of factors like age, gender, race, or social status. Finally, to foster public acceptance, manufacturers and policymakers must address concerns that may discourage potential AV buyers about their decision to adopt AV technology (Rafiee et al., 2023).

The outlined objectives serve as a foundational guide for policymakers and manufacturers in navigating moral dilemmas related to potential human life loss in autonomous cars. However, the ethical landscape of AVs is complex and continually evolving. Policymakers must recognise the need for additional objectives, and their selection should be approached with caution, given potential conflicts. Public engagement and interdisciplinary collaboration involving ethicists, sociologists, psychologists, and technologists are crucial for a comprehensive understanding of ethical implications (Jobin et al., 2019). Value conflicts may arise, necessitating transparent decision-making processes to address trade-offs, demographic prioritisation, and algorithmic transparency. The chosen objectives should reflect a balanced approach, acknowledging tensions and striving for ethically sound solutions. A commitment to responsible development, ongoing dialogue, and adaptability will contribute to a robust ethical foundation, fostering public trust and aligning AV technology with societal values (Amodei et al., 2016).

3.2 Argument-2 “Existing Solution to AV Moral Challenges”

The NTC's second reason for the exclusion of ethical principles criterion in the DRIS, was influenced by a proposed solution for AVs in dealing with moral dilemmas. The proposal originated from a conference paper by two Swedish academics, which suggests that if an ADS can assess its operational capacity to handle unexpected situations and adjust its behaviour accordingly, the probability of facing ethical dilemmas will be relatively low (Johansson and Nilsson, 2016). In our assessment we will examine the assumptions that were used in the paper and evaluate the arguments in order to determine their soundness.

3.2.1 Irrelevance of Trolley like Scenarios

Johansson and Nilsson first argue that other publications regard the “Trolley Problem” like thought-experiments as a real and important issue to address in the context of AVs and not as a constructed problem used to discuss ethical dilemmas. Through this assertion, their aim is to demonstrate that such events are rare or nearly impossible, deeming them irrelevant (Johansson and Nilsson, 2016). The “Trolley Prob-

lem” is a classic ethical dilemma involving a runaway trolley on a track about to harm and kill five people. You have the choice to either divert the trolley to a side track by pulling a lever, causing harm to one person, or do nothing and let it harm the five people on the main track. The determination of the “correct” decision remains a subject of ongoing philosophical debate, as both choices appear reasonable and defensible. It is essential to clarify that the value lies not in arriving at a definitive answer but rather to cultivate the ability to engage in meaningful debates for either option. This perspective is particularly highlighted by the different variations explored in Thomson’s 2019 study, emphasising the complexities inherent in such ethical dilemmas (Thomson, 2019).

Although the “Trolley Problem” might appear abstract and even fantastical due to the simplicity of its setup and the intricacies of its variations, it is deemed a valuable representation of moral dilemmas. Undoubtedly AVs will encounter complex moral situations that cannot be overlooked. With specific regard to AVs, the notion of a singular, precisely defined moral dilemma may be deemed highly unlikely. ADS may have the ability to optimise their behavior over a time horizon, minimising the likelihood of facing a singular, immediate moral dilemma. It might be reasonable to argue that ASDs should then prioritise minimizing the risk of encountering such dilemmas in the future. For instance, in a moral dilemma involving an AV, the vehicle could proactively slow down to provide more reaction time, signal to nearby vehicles to maintain a safer distance, or prompt a motorcyclist to either keep a greater distance or pass the vehicle safely. Regardless, the fact remains both taking action and non-action can lead to severe consequences for human well-being. In such cases, a thoughtful evaluation of the harm imposed, often affecting innocent third parties not initially part of the situation, is crucial when comparing the potential loss of life resulting from non-action or default action.

3.2.2 Vienna Convention

The Vienna Convention (VC) on Road Traffic, established in 1968, serves as an international treaty with the goal of harmonising rules and regulations for road traffic among its participating nations. The convention aims to facilitate international road traffic and improve road safety by standardising traffic regulations. Citing Article 8 of the VC, the authors emphasise the significance of drivers maintaining control over their vehicles, underlining the driver’s responsibility to effectively manage their vehicles for overall road safety (UNECE, 1968). The authors contend that this argument implies drivers should possess the fore-

sight to anticipate dilemma situations and adjust their driving behavior accordingly. They propose a shift in focus from addressing ethical dilemmas to educating drivers on accident avoidance. Emphasising the societal responsibility placed on drivers, they point out that obtaining a driver’s license requires demonstrating the ability to control the vehicle, even in unforeseen events (Johansson and Nilsson, 2016). They present a scenario where a child unexpectedly appears in front of the car, highlighting that the driver cannot hold the child responsible. They suggest that it is the driver’s duty to anticipate such situations through foresight and adjust their driving behaviour accordingly. Their recommended approach involves directing attention to implementing robust safety measures in AVs to proactively prevent such scenarios.

Firstly, the challenge individuals encounter in consistently behaving ethically, even after learning correct conduct, arises from a complex interplay of psychological, behavioral economic, and ethical factors. Psychological factors such as cognitive dissonance and moral licensing play crucial roles. Cognitive dissonance emerges when personal beliefs clash with actions, leading individuals to justify unethical behavior. Moral licensing occurs when prior ethical actions permit subsequent unethical conduct (Schwartz, 2008; Kahneman, 2011; Bandura, 2014). From a behavioral economics standpoint, individuals often display predictably irrational behavior influenced by cognitive biases (Kahneman, 2011). Time discounting results in people prioritising immediate gains over long-term ethical considerations. Ethical decision-making is further complicated by variations in moral development and situational ethics. Moral development influences an individual’s capacity to prioritise ethical considerations over personal interests, while situational factors affect the consistent application of ethical principles across diverse scenarios (Kohlberg, 1973; Bandura, 2014). Additionally, making thoughtful ethical decisions involving various alternatives with specific stakes in unfamiliar situations requires time, focus, and careful consideration. In emergency situations, drivers lack these resources, relying on instinct and intuition, often in a surprised and panic-stricken manner. Ethics training offers no benefit in such scenarios, resulting in zero net effect. However, the programming of AVs may influence their performance in ethical dilemmas due to their capabilities.

Secondly, we dispute the authors’ claim that “dealing with dilemma situations are not covered in any driver instruction textbook, as the primary focus has been on education drivers to avoid accidents”. A more precise statement might be that spe-

cialised training specifically designed to prepare human drivers for such situations is not mandatory. At least in Australia, safe driving courses are available and play a crucial role in educating drivers about responsible and safe driving practices, covering both theoretical knowledge and practical skills. These courses provide instruction in defensive driving techniques, equipping participants with strategies to anticipate and effectively respond to potential road hazards (PCYC, 2023; Sense, 2023).

Finally, we could draw a fitting analogy between the programming of AVs to navigate ethical dilemma situations and the concept of insurance. Much like individuals choose car or home insurance to safeguard their assets, even if they never end up filing a claim, the proactive preparation of AVs for ethical decision-making can be seen as a form of assurance. While AVs are anticipated to be significantly safer than human drivers, the foresighted consideration of how they would handle moral dilemmas is similar to an insurance policy, providing a layer of preparedness for unforeseen situations.

3.2.3 Core Safety Proposition

The authors make a reference to the ISO26262 terminology, which is an international standard for functional safety of electrical and electronic systems within the automotive industry. It provides guidelines and requirements to ensure the safety of automotive systems, focusing on the prevention and control of hardware and software failures that could lead to hazardous situations. The standard is applicable to the entire life-cycle of a vehicle, from concept to decommissioning (Hommes, 2012). They explain that, in the context of functional safety, the risk is measured by taking into account the severity and frequency of a hazardous event. If the event's severity is very low or the likelihood of a hazardous event is deemed highly improbable (E0), there is no requirement for a specific risk measure, and in the context of AV, ethical considerations can be disregarded. The authors proceed with the assumption that once all hazardous events, even those with minimal probability, are taken into account, it can be asserted that the ADS functionality behaves safely (Johansson and Nilsson, 2016).

The authors suggest that operational algorithms play a crucial role in determining tactical moves in driving scenarios. Using the example of an overtaking manoeuvre, the operational capability of a system is key in deciding whether a tactical move is executed. If the operational algorithm is confident in its capabilities, it will implement the manoeuvre. However, if a task is considered "risky", it should be avoided. The level of operational capability influences the driving

style, with higher capability allowing for a "normal" style and lower capability resulting in a more conservative, but always safe, driving approach (Johansson and Nilsson, 2016).

Their argument is primarily based on the idea that all potential risks can be identified in advance, a notion we previously discussed as nearly impossible. Additionally, we clarified that the progress of AV technology will occur gradually, resembling the evolution seen in commercial airliners, meaning that it will not attain perfection overnight. This highlights the importance of ethical considerations in the development and deployment of AV technology.

Even if we assume that an ADS could anticipate every hazardous scenario beforehand, there are situations where it would have to limit its tactical choices to match the operational capability necessary for planning the most unexpected events. This could reduce its operational capability to the extent that coming to a stop would be deemed the safest tactical choice for a particular scenario. Such unpredictable starts and stops would not only perplex other ADSs but also manual drivers and other road users, including pedestrians, who would not be able to anticipate the AV's actions due to their radical nature. In other words, a default policy of "just stop" can be a risky and unpredictable default response, compared to more sophisticated ethical actions.

4 RECOMMENDATION FOR AUSTRALIAN POLICYMAKERS

An immediate and widespread deployment of fully automated vehicles appears unlikely, particularly in the near future. Instead, the progression of AVs is anticipated to follow a gradual trajectory, evolving from controlled and specialised conditions (operational design domains) to more complex scenarios. This gradual approach, coupled with extensive planning, regulatory considerations, and infrastructure development, carries significant implications for how the government should anticipate and handle these transformative changes (iMOVE, 2023). In this section, we have compiled a set of recommendations specifically tailored for the Australian government and policymakers, with the goal of enhancing the safety of ADSs and fostering greater acceptance within the community.

4.1 Regulation

Traditionally, governments have been responsible for overseeing road and vehicle safety, and community feedback consistently indicates an expectation for government involvement in ensuring the safety of AVs (NTC, 2017). Given the novelty of these vehicles and the uncertainty surrounding their safety performance, government engagement offers reassurance and clarity to industry, insurers, and consumers, establishing a foundational safety standard as AVs become more common.

The absence of clear regulations may leave governments without the means to ensure safety, causing uncertainty among various stakeholders. Research suggests the necessity for a delicate regulatory approach that balances two conflicting goals. First, it aims to protect public safety from potential hazards arising from underdeveloped and inadequately engineered ADSs that may lead to accidents. Second, it aims to promote technological advancements in vehicle technology for future safety improvements (Nowakowski et al., 2015). Although the government cannot directly dictate how car manufacturers program their ADS, they can establish ethical rules and guidelines prioritising safe testing over technological advancements for public safety (iMOVE, 2023).

4.2 Educate and Inform

The government can play a crucial role in fostering public acceptance of AVs by implementing strategic educational initiatives. Collaborating with community leaders and early adopters from diverse backgrounds can serve as a catalyst for broader acceptance within the community. Public information campaigns, highlighting ongoing and upcoming trials and deployments, can be instrumental in disseminating key information. Emphasis should be placed on communicating the benefits of ADS, such as enhanced mobility and accessibility, while addressing prevalent concerns related to safety, costs, and liability. Tackling these critical issues head-on can serve as a pivotal factor in fostering greater acceptance (Orr, 2003).

Additionally, trust in new technology often grows with increased exposure and repeated use. Encouraging pilot programs and demonstrations that allow the public to interact with ADS in controlled environments can contribute significantly to building trust. By proactively addressing concerns, showcasing benefits, and involving community influencers, the government can actively contribute to the successful integration and acceptance of AV technology.

4.3 The Ethical Dimension

The development and deployment of AVs is intrinsically linked to both safety and ethics. While AVs hold the potential to enhance road safety, their introduction poses ethical challenges related to decision-making in complex scenarios. The safety risk of AVs cannot be adequately managed unless ethical considerations are incorporated into discussions and the development of these technologies.

The interplay between ethics and safety is evident in the need to ensure that AVs are programmed to prioritise safety while making ethically sound decisions. Developers, regulators, and the wider public must be engaged in discussions about the ethical considerations that shape AV behavior. The decision-making process should be transparent and subject to public scrutiny. The programming of these algorithms should ensure that ethical choices are in harmony with the values upheld by society.

As we previously discussed, AVs will be equipped with algorithms that determine how they respond to various scenarios. Their design should focus on accident avoidance whenever feasible. In the exceptional instances where an accident is unavoidable, AVs need to be programmed to decide how harm is allocated among all parties involved, whether they are passengers, pedestrians, or other road users.

5 CLOSING REMARKS

While the NTC's recommended safety criteria offer a foundation for addressing safety dilemmas that carry ethical implications, they remain at a high level and lack specific guidance for ADSEs in navigating the intricate task of programming their ADS with considerations for the mentioned principles. The core idea behind the safety criteria is that every AV should operate within its predefined operational design or be capable of transitioning into a minimal risk condition in emergency situations. However, we have demonstrated that this might not always be feasible, as there are instances where the ADS might need to make difficult choices that could result in significant injury or even the loss of human life.

The German Ethics Commission on Automated Driving, established by the Federal Ministry of Transport and Digital Infrastructure, addresses the ethical considerations related to the programming of AVs. One of the key principles endorsed by the commission opposes the programming of AVs to make decisions between human lives, emphasising a commitment to avoiding scenarios where AV program-

ming intentionally prioritises one life over another (Luetge, 2017). However, the NTC, while acknowledging the German Ethics Commission's stance, rejects the necessity for such considerations and discussions, asserting that programming decisions involving human lives are deemed impractical under any circumstances. We recognise the rarity of dilemma situations and the current state of AV technology, and therefore understand that prioritising one road user over another for the sake of equal treatment may not be feasible. However, as mentioned earlier, these inevitable situations are likely to be encountered by an ADS eventually. Therefore, we assert that the inclusion of ethical considerations for AVs, especially in complex ethical dilemmas, should be part of future discussions, even in the absence of a clear consensus on ethical decision-making by ADSs.

REFERENCES

- Amodei, D., Olah, C., Steinhardt, J., Christiano, P., Schulman, J., and Mané, D. (2016). Concrete problems in ai safety. *arXiv preprint arXiv:1606.06565*.
- Bandura, A. (2014). Social cognitive theory of moral thought and action. In *Handbook of moral behavior and development*, pages 69–128. Psychology press.
- Coca-Vila, I. (2018). Self-driving cars in dilemmatic situations: An approach based on the theory of justification in criminal law. *Criminal Law and Philosophy*, 12(1):59–82.
- Dewilde, M. (2022). A comparative study between aerospace manufacturers boeing and airbus.
- FCAI (2023). FCAI submission to NTC discussion paper on control of automated vehicles.
- Hommel, Q. V. E. (2012). Review and assessment of the iso 26262 draft road vehicle-functional safety. Technical report, SAE Technical Paper.
- iMOVE (2023). Promoting community readiness and uptake of cavs. Online.
- Jobin, A., Ienca, M., and Vayena, E. (2019). The global landscape of ai ethics guidelines. *Nature machine intelligence*, 1(9):389–399.
- Johansson, R. and Nilsson, J. (2016). Disarming the trolley problem—why self-driving cars do not need to choose whom to kill. In *Workshop CARS 2016-Critical Automotive applications: Robustness & Safety*.
- Kahneman, D. (2011). *Thinking, fast and slow*. macmillan.
- Kira, J. (2017). Consultant report: Safety assurance system for automated vehicles in Australia.
- Kohlberg, L. (1973). The claim to moral adequacy of a highest stage of moral judgment. *The journal of philosophy*, 70(18):630–646.
- Krueger, R., Rashidi, T. H., and Rose, J. M. (2016). Preferences for shared autonomous vehicles. *Transportation research part C: emerging technologies*, 69:343–355.
- Luetge, C. (2017). The german ethics code for automated and connected driving. *Philosophy & Technology*, 30:547–558.
- Maurer, M., Gerdes, J. C., Lenz, B., and Winner, H. (2016). *Autonomous driving: technical, legal and social aspects*. Springer Nature.
- Nowakowski, C., Shladover, S. E., Chan, C.-Y., and Tan, H.-S. (2015). Development of california regulations to govern testing and operation of automated driving systems. *Transportation Research Record*, 2489(1):137–144.
- Noy, I. Y., Shinar, D., and Horrey, W. J. (2018). Automated driving: Safety blind spots. *Safety Science*, 102:68–78.
- NTC (2017). Regulatory options to assure automated vehicle safety in Australia: discussion paper.
- NTC (2018). Safety assurance for automated driving systems decision regulation impact statement.
- NTC (2022). Regulatory framework for automated vehicles in Australia.
- Nyholm, S. and Smids, J. (2016). The ethics of accident-algorithms for self-driving cars: An applied trolley problem? *Ethical theory and moral practice*, 19(5):1275–1289.
- Orr, G. (2003). Diffusion of innovations, by everett rogers (1995). Retrieved January, 21:2005.
- PCYC (2023). Pyc driver education.
- Rafiee, A., Wu, Y., and Sattar, A. (2023). Philosophical and legal approach to moral settings in autonomous vehicles: An evaluation. In *Social Licence and Ethical Practice*, volume 27, pages 95–114. Emerald Publishing Limited.
- Schwartz, H. (2008). Predictably irrational: the hidden forces that shape our decisions. *Business Economics*, 43(4):69–72.
- Sense, R. (2023). Road sense driver education course.
- Shadrin, S. S. and Ivanova, A. A. (2019). Analytical review of standard sae j3016 "taxonomy and definitions for terms related to driving automation systems for on-road motor vehicles" with latest updates. *Avtomobil'. Doroga. Infrastruktura.*, (3 (21)):10.
- Thomson, J. J. (2019). Killing, letting die, and the trolley problem. In *Death, Dying and the Ending of Life, Volumes I and II*, pages V2.17–V2.30. Routledge.
- UNECE (1968). Convention on road traffic done at vienna on 8 november 1968.
- Wu, S. S. (2020). Autonomous vehicles, trolley problems, and the law. *Ethics and Information Technology*, 22(1):1–13.
- Yadron, D. and Tynan, D. (2016). Tesla driver dies in first fatal crash while using autopilot mode. *The Guardian*.

APPENDIX

Safety Criteria for Automated Driving Systems

The ADSE, rather than government, will be responsible for testing and validating the safety of the ADS or function and documenting these processes. The role of government is to satisfy itself that the applicant has processes in place to identify and manage the safety risks. This broad approach was agreed to by transport ministers in November 2017, subject to analysis through a Regulation Impact Statement (NTC, 2018).

Eleven safety criteria that require the applicant to demonstrate its processes for managing safety risks:

1. Safe system design and validation processes
2. Operational design domain (ODD)
3. Human-machine interface (HMI)
4. Compliance with relevant road traffic laws
5. Interaction with enforcement and other emergency services
6. Minimal risk condition
7. On-road behavioural competency
8. Installation of system upgrades
9. Verifying for the Australian road environment
10. Cybersecurity
11. Education and training.