RESEARCH REPORT

Institute of Transportation Studies | UNIVERSITY OF CALIFORNIA

A Blueprint for Improving Automated Driving System Safety

Mollie Cohen D'Agostino, Executive Director, Mobility Science Automation and Inclusion Center (MoSAIC), Institute of Transportation Studies, University of California, Davis
Cooper Michael, J.D., Graduate of the University of California Davis School of Law
Marilia Ramos, Former Research Scientist, B. John Garrick Institute for The Risk Sciences, University of California Los Angeles
Camila Correa-Jullian, Graduate Student Researcher, B. John Garrick Institute for The Risk Sciences, University of California Los Angeles

July 2024

California Resilient and Innovative Mobility Initiative

Report No.: UC-ITS-RIMI-5B-02 | DOI: 10.7922/G2RR1WKG

Technical Report Documentation Page

1. Report No. RIMI-5B-02	**2. Government Accession No.** N/A	**3. Recipient's Catalog No.** N/A
4. Title and Subtitle A Blueprint for Improving Automated Driving System Safety		**5. Report Date** July 2024
		6. Performing Organization Code ITS-Davis
7. Author(s) Mollie Cohen D'Agostino, https://orcid.org/0000-0002-3689-9471 Cooper Elliott Michael, https://orcid.org/0009-0007-6254-5339 Marilia Ramos, https://orcid.org/0000-0002-5039-5747 Camila Correa-Jullian https://orcid.org/0000-0002-4622-4064		**8. Performing Organization Report No.** N/A
9. Performing Organization Name and Address Institute of Transportation Studies, Davis 1605 Tilia Street Davis, CA 95616		**10. Work Unit No.** N/A
		11. Contract or Grant No. RIMI-5B-02
12. Sponsoring Agency Name and Address The University of California Institute of Transportation Studies www.ucits.org		**13. Type of Report and Period Covered** White Paper (November 2022 – October 2023)
		14. Sponsoring Agency Code UC ITS
15. Supplementary Notes DOI: 10.7922/G2RR1WKG		

16. Abstract

Vehicle automation represents a new safety frontier that may necessitate a repositioning of our safety oversight systems. This white paper serves as a primer on the technical and legal landscape of automated driving system (ADS) safety. It introduces the latest AI and machine learning techniques that enable ADS functionality. The paper also explores the definitions of safety from the perspectives of standards-setting organizations, federal and state regulations, and legal disciplines. The paper identifies key policy options building on topics raised in the White House's Blueprint for an AI Bill of Rights, outlining a Blueprint for ADS safety. The analysis concludes that potential ADS safety reforms might include either reform of the Federal Motor Vehicle Safety Standards (FMVSS), or a more holistic risk analysis "safety case" approach. The analysis also looks at caselaw on liability in robotics, as well as judicial activity on consumer and commercial privacy, recognizing that the era of AI will reshape liability frameworks, and data collection must carefully consider how to build in accountability and protect the privacy of consumers and organizations. Lastly, this analysis highlights the need for policies addressing human-machine interaction issues, focusing on guidelines for safety drivers and remote operators. In conclusion, this paper reflects on the need for collaboration among engineers, policy experts, and legal scholars to develop a comprehensive Blueprint for ADS safety and highlights opportunities for future research.

17. Key Words Automated vehicle control, traffic safety, case law, policy, machine learning, artificial intelligence		**18. Distribution Statement** No restrictions.	
19. Security Classification (of this report) Unclassified	**20. Security Classification (of this page)** Unclassified	**21. No. of Pages** 57	**22. Price** N/A

Form Dot F 1700.7 (8-72) Reproduction of completed page authorized

About the UC Institute of Transportation Studies

The University of California Institute of Transportation Studies (UC ITS) is a network of faculty, research and administrative staff, and students dedicated to advancing the state of the art in transportation engineering, planning, and policy for the people of California. Established by the Legislature in 1947, ITS has branches at UC Berkeley, UC Davis, UC Irvine, and UCLA.

The California Resilient and Innovative Mobility Initiative

The California Resilient and Innovative Mobility Initiative (RIMI) serves as a living laboratory – bringing together university experts from across the four UC ITS campuses, policymakers, public agencies, industry stakeholders, and community leaders – to inform the state transportation system's immediate COVID-19 response and recovery needs, while establishing a long-term vision and pathway for directing innovative mobility to develop sustainable and resilient transportation in California. RIMI is organized around three core research pillars: Carbon Neutral Transportation, Emerging Transportation Technology, and Public Transit and Shared Mobility. Equity and high-road jobs serve as cross-cutting themes that are integrated across the three pillars.

Acknowledgments

This study was made possible with funding received by the University of California Institute of Transportation Studies from the State of California through a one-time General Fund allocation in the 2021 State Budget Act for the Resilient and Innovative Mobility Initiative. The authors would like to thank the State of California for its support of university-based research, and especially for the funding received for this project.

Disclaimer

Institute of Transportation Studies | UNIVERSITY OF CALIFORNIA

A Blueprint for Improving Automated Driving System Safety

Mollie Cohen D'Agostino, Executive Director, Mobility Science Automation and Inclusion Center (MoSAIC), Institute of Transportation Studies, University of California, Davis
Cooper Michael, J.D., Graduate of the University of California Davis School of Law
Marilia Ramos, Former Research Scientist, B. John Garrick Institute for The Risk Sciences, University of California Los Angeles
Camila Correa-Jullian, Graduate Student Researcher, B. John Garrick Institute for The Risk Sciences, University of California Los Angeles

July 2024

California Resilient and Innovative Mobility Initiative

Report No.: UC-ITS-RIMI-5B-02 | DOI: 10.7922/G2RR1WKG

Table of Contents

Table of Contents

List of Figures

List of Tables

Acronyms

ADS	automated driving system
ADAS	automated driving assistance system
ADS-DV	automated driving system–dedicated vehicle
ANPRM	advance notice of proposed rulemaking
API	application programming interface
AV	automated vehicle
CAV	connected automated vehicle
CPUC	California Public Utilities Commission
DSRC	dedicated short-range communications
FAA	Federal Aviation Administration
FCC	Federal Communication Commission
FMVSS	Federal Motor Vehicle Safety Standards
FSD	full self-driving
MDS	mobility data specification
NHTSA	National Highway Traffic Safety Administration
NTSB	National Transportation Safety Board
ODD	operational design domain
QRN	quantitative risk norm
SOTIF	safety of the intended functionality

Glossary

Automated Driving Systems (ADS)	Systems that can perform all or part of driving tasks on a sustained basis (ISO 34501:2022) ; "set of elements (3.14) that offer a specific conditional or higher automated driving use case (3.63) in or for a specific operation design domain (ODD) (3.37)." (ISO/TR 4204:2020(en)
ADS Safety Policy Framework	Comprehensive system of regulatory policy that establishes oversight protocols, timelines, and clearly defined roles that align with public policy goals for safety. This type of framework would ensure that the ADS sector realizes safety goals.
Automation Complacency	Vehicle operators (e.g., safety driver or remote operator) build excessive trust in a vehicle's good performance and fail to monitor it effectively.
Consumer-Expectations Test	Test to determine whether a product did not perform as safely as an ordinary consumer would expect when used in the intended or reasonably foreseeable manner.
Deep Reinforcement Learning	(See Reinforcement learning) Neural networks of reinforcement learning models that can determine, using multiple data inputs, the safest trajectory path for ADS travel.
Fatigue Risk Management Policies	Procedures in a safety case that manage safety risks related to driver fatigue.
Hazard Analysis and Risk Assessment (HARA)	Per ISO, an assessment of hazardous events or a combination of hazards and an operational situation.
Human Out-of-the-Loop Issues	Issues where humans lack context to take rapid actions for hazard avoidance.
Imitation Learning	A machine-learning strategy that trains ADS using human behavior as a guide.
Lagging Metrics	Metrics that report observable failure events, like accidents, injuries, fatalities.
Leading Metrics	An umbrella term for nearly all issues that do not result in a catastrophic failure event, like a vehicle collision. These include hard breaking incidents, near-misses, unplanned stops, or instances of vehicles blocking roadways, traffic zones, or driveways.

Minimal Risk Condition (MRC)	State a vehicle achieves or aims to achieve to mitigate the risk of an incident after it has occurred (*usually a stable, stopped condition in a safe location as context permits).
Operational Design Domain (ODD)	Operating conditions under which a given driving automation system or feature thereof is specifically designed to function, including, but not limited to, environmental, geographical, and time-of-day restrictions, and/or the requisite presence or absence of certain traffic or roadway characteristics.
Reinforcement Learning	Reinforcement learning is a type of machine learning approach that is used to train the ADS, and based on the learned policy, performs decision-making tasks in real time.
Risk-Utility Test	Judicial test that weighs the danger of the vehicle against its usefulness.
Safe State	When the system can "avoid risk, in an acceptable criterion, to any road user" (ISO 4804:2020(en)
Safety Case	A structured argument, supported by evidence, to justify that a system is acceptably safe for a specific application in a specific operating environment.
Safety of the Intended Functionality (SOTIF)	SOTIF is the absence of unreasonable risk resulting from hazardous behaviors related to function insufficiencies (ISO). It is also defined as the absence of unreasonable risk due to hazards resulting from functional insufficiencies of the intended functionality or from reasonably foreseeable misuse by persons (British Standards Institute).

Executive Summary

Executive Summary

Vehicle automation is introducing a new motor vehicle safety frontier. Driving automation may reshape traffic safety norms and require a repositioning of our safety oversight systems.[1] [2] This white paper intends to be a primer on the technical and legal landscape of automated driving system (ADS) safety. The paper introduces the latest in AI and machine learning techniques that make ADS possible (Section I). The Background section (Section II) explores multidisciplinary definitions of safety through three lenses, standards-setting organizations, among federal and state regulations, and within the legal discipline. Section III identifies key policy considerations to inform guardrails for the safety of ADS systems.

Section I: What are Automated Driving Systems?

Automated driving technology is currently classified through six Levels of Driving Automation, ranging from driver assistance (L0-L2) to automated driving (L3-L5) functionalities. The term ADS is used to refer to systems from Level 3 onwards, at which the vehicle can perform limited driving tasks under certain conditions. As the level of driving automation increases, the driving tasks are progressively transferred from the driver to the vehicle, until little to no human intervention is expected to be required. ADS uses a combination of sensors to perform real-time perception of the environment, categorize other vehicles, pedestrians, or objects to inform trajectory planning for determining a vehicle's next movement.[3] These sensors may include cameras, radar, lidar, inertial sensors, and GPS. The subsequent localization, path planning and decision-making tasks can rely on a combination of sensor fusion strategies, rule-based algorithms, high-definition (HD) mapping, and neural networks architectures. Recent advances in neural networks, including reinforcement learning, and foundation world models, are improving the accuracy of these decisions. However, concerns abound about the opacity of these deep learning algorithms, and the unpredictability inherent in systems of continuous learning will make establishing safety standards an ongoing challenge.[4] [5]

[1] AUTOMATED VEHICLES - COMPREHENSIVE PLAN, (2021), HTTPS://WWW.TRANSPORTATION.GOV/AV/AVCP.
[2] DASOM LEE & DAVID J. HESS, REGULATIONS FOR ON-ROAD TESTING OF CONNECTED AND AUTOMATED VEHICLES: ASSESSING THE POTENTIAL FOR GLOBAL SAFETY HARMONIZATION, 136 TRANSPORTATION RESEARCH PART A: POLICY AND PRACTICE 85 (2020).
[3] EMMANUEL OWUSU APPIAH & SOLOMON MENSAH, *OBJECT DETECTION IN ADVERSE WEATHER CONDITION FOR AUTONOMOUS VEHICLES*, 83 MU LTIMEDIA TOOLS AND APPLICATIONS 28235 (2024).
[4] DAVID F. LLORCA ET AL., *LIABILITY REGIMES IN THE AGE OF AI: A USE-CASE DRIVEN ANALYSIS OF THE BURDEN OF PROOF*, 76 JOURNAL OF ARTIFICIAL INTELLIGENCE RESEARCH 613 (2023).
[5] JAN DE BRUYNE, ELIAS VAN GOOL & THOMAS GILS, *TORT LAW AND DAMAGE CAUSED BY AI SYSTEMS*, *IN* ARTIFICIAL INTELLIGENCE AND THE LAW 359 (JAN DE BRUYNE & CEDRICEDITORS VANLEENHOVE EDS., 2021).

Section II: Defining Safety

This paper begins by investigating the fundamentals of safety for ADSs. Definitions of *safety* vary across different disciplines. *Safety* is not a singular concept, but a multitude of concepts. Safety is said to reflect cultural norms, and *safety culture* can be an organizational tool to advance safe decision-making.[6] [7] Safety can be viewed as a measurement, a process, or a threshold. ADS safety experts introduce the concept of *roadsmanship*, which means driving in a way that responds safely to hazards and avoids creating hazards for others. Defining what constitutes safe ADS vehicle performance may include the development of new vehicle standards, use of human reference standards, and new methods for evaluating risk.

Risk underpins the definitions for safety for organizations like the Society of Automotive Engineers (SAE) and the International Organization for Standardization (ISO). These standard-setting organizations developed consensus around how to define safety and assess risk. The ISO defines *safety* as the "absence of unreasonable risk", and many other organizations maintain similar language. Standards-setting organizations are ahead of regulators in adding more specificity to ADS engineering concepts and establishing protocols, and several industry-wide norms for ADS operation have emerged. Leading standards for ADS safety include Underwrite Laboratories (UL) 4600. UL 4600 is a standard that suggests hundreds of possible known faults and hazards, as well as strategies for evaluating and mitigating risks related to these factors. The goal of following this standard is to demonstrate an absence of unreasonable risk.

This concept of unreasonable risk is codified in U.S. federal statute (49 U.S.C. § 30102) where regulators define "motor vehicle safety" as protecting the "public against unreasonable risk of accidents occurring because of the design, construction, or performance of a motor vehicle…" Manufacturers must comply with U.S. federal motor vehicle safety standards (FMVSS) and certify their vehicles achieve this standard of safety before operating. These standards are largely self-administered, and federal regulators retain only a limited and reactive recall authority, collecting empirical data to monitor safety, and suggesting recalls when evidence of unreasonable risks emerges. States, like California, have somewhat more robust evaluation procedures for assessing ADS safety performance. States also have clear jurisdiction to focus on passenger safety in commercial operations, like taxis, ridehail, or private shuttles.

Courts and legal scholars are interpreting safety with an emphasis on liability. Legal precedent is still emerging, and this paper describes several historical cases involving liability in robotics for manufacturing, medical surgery settings, street-cleaning robots, and other AI-enabled consumer electronics products. These accidents raise fascinating questions about how to assess faults and damages in AI systems. Legal ethicists point to the role of *duty of care* in the judicial evaluation of liability for traffic collisions, where legal precedent supposes drivers must abide by a social contract to take reasonable and prudent actions, even when doing so would

[6] MARJORY S. BLUMENTHAL ET AL., *SAFE ENOUGH: APPROACHES TO ASSESSING ACCEPTABLE SAFETY FOR AUTOMATED VEHICLES*, (2020), HTTPS://WWW.RAND.ORG/PUBS/RESEARCH_REPORTS/RRA569-1.HTML (LAST VISITED APR 16, 2023).
[7] K. HAUKELID, THEORIES OF (SAFETY) CULTURE REVISITED—AN ANTHROPOLOGICAL APPROACH, 46 SAFETY SCIENCE 413 (2008).

violate traffic code, if doing so would avoid catastrophic outcomes.[8] [9] This legal concept of *duty of care*, when applied to ADS, aligns with the engineering concept of *roadmanship*.

Section III. ADS Policy Considerations

The White House released a *Blueprint for an AI Bill of Rights* in 2022 identifying several topics that are highly relevant to ADS safety. We apply several of these topics as a *blueprint* for our ADS policy analysis, focusing on 1) safety, 2) data collection and privacy, and 3) human-machine interfacing issues.

Blueprint for General Safety Policy: Two approaches are identified as options for advancing the overarching strategy of ADS safety policy. Further investigation will be necessary to determine which combinations of these represent the best path forward.

- **Reform of FMVSS:** This approach is relegated to federal regulatory actions. Reforms might include adding new FMVSSs for ADS vehicles, reforming exemption caps and processes. Reforms might also include building institutional investments in simulation platforms, or creating a standardized scenario catalog for simulation, or developing a human standard for simulation comparison, or for ongoing fleet operational safety analyses.
- **Safety Case Approach:** This approach could be implemented at state or federal levels. It would establish a pathway for ADS manufacturers to develop and maintain a *Safety Case* document (in lieu of complying with certain FMVSS, or in addition). A Safety Case is a list of safety claims, with supporting evidence. The Main Case, or argument, is divided into subclaims demonstrating through detailed evidence, and then each case can be further subdivided, until all known hazards can be shown to be recognized and mitigated (and limitations or unknowns explicitly stated). Independent oversight bodies evaluate Safety Case submissions in other industries (e.g., nuclear, maritime) and make compliance recommendations to decision-makers. A Safety Case could be created for initial demonstration of minimum viability and/or updated regularly based on triggering events (e.g., fleet scaling milestones, new land-use types, etc.).[10]

Blueprint for ADS Data Policy: The dependence of ADS-equipped vehicles on external sensors to understand the world and make informed decisions requires gathering massive amounts of data. Regulators will have to thread a needle on what data is necessary and whether data collection can have a clear use case that serves explicit policy goals. Modernizing and aligning data formats will also streamline data collection. These protocols must also be developed with the privacy interests of consumers and proprietary interests in mind.

---

[8] A D'AMATO ET AL., EXCEPTIONAL DRIVING PRINCIPLES FOR AUTONOMOUS VEHICLES, MOB. 2 UNIV. MICH. J. LAW MOBIL., HTTPS://FUTURIST.LAW.UMICH.EDU/EXCEPTIONAL-DRIVING-PRINCIPLES-FOR-AUTONOMOUS-VEHICLES/.

[9] J. CHRISTIAN GERDES & SARAH M. THORNTON, *IMPLEMENTABLE ETHICS FOR AUTONOMOUS VEHICLES, IN AUTONOMOUS DRIVING: TECHNICAL, LEGAL AND SOCIAL ASPECTS* 87 (MARKUS MAURER ET AL. EDS., 2016), HTTPS://DOI.ORG/10.1007/978-3-662-48847-8_5.

[10] THOR MYKLEBUST, FUTURE CHALLENGES, PITFALLS, AND OPPORTUNITIES WHEN USING A SAFETY CASE APPROACH FOR SW-INTENSIVE SYSTEMS, (2023).

General Data Collection Policy: Ongoing oversight and additional guidance will ensure that each ADS fleet can scale safely, grow competence, and demonstrate continuous improvement. Duplicative data collection is occurring across regulatory bodies, and a shared API or common data specification might improve compatibility and enable a centralized data collection system where states can easily add relevant criteria. The paper highlights several data collection examples that may serve as a model for ADS data collection, including the USDOT Data exchange, the mobility data standard (MDS), or the Federal Aviation Administration (FAA)'s Aviation Safety Information and Sharing (ASIAS) program. These examples might address different data needs across the following governmental authorities:

- Federal ADS data exchange could evaluate ADS performance, overseen by an independent evaluation body, to collect: 1) Leading metrics (e.g., near-misses, minimal risk condition events), and 2) Lagging metrics, (e.g., observable failure events, accidents, injuries, and fatalities).
- State data collection strategies might focus on assuring vehicle safety meets state standards and will focus on passenger service vehicles (e.g., taxi, ridehailing, shuttles). This will include a focus on consumer safety, sustainability, equity, as well as other topics that align with state policy priorities. States might also focus data collection on workforce tracking metrics.

ADS Data Privacy Considerations: Collecting ADS data is not without risks and tradeoffs. Movement and routing data, video, or images of individuals inside the vehicles, as well as personal information regarding riders age, gender, or other characteristics [11] means that many ADS collected data are classified as sensitive personally identifiable information (PII). Purview over privacy laws is shared by federal authorities, states, and other jurisdictions, and more coordination on the following types of policies will be key to ensure that privacy is preserved across both consumers and commercial interests.

- **Consumer privacy considerations:** Privacy risk assessments can help ensure accountability of all stakeholders (these could be a new FMVSS or an element of a Safety Case). These assessments could include independent tests for data security, to ensure consumer data is held securely and held discrete from personal demographic information in alignment with regulatory requirements. For example, the Privacy Act of 1974 offers some protections on retention of personal information, "including limits on data retention, and also provides individuals a general right to access and correct their data." [12]
- **Proprietary Interest Privacy Considerations:** Establishing clear criteria for what data constitutes a trade secret, and setting clear limitations on why this information can be collected as well as who can view raw data in this category, will help preserve the integrity of any trade secret information where it is determined to be necessary to collect this type of information.

Blueprint for Policy Regarding Human Alternatives and Fallback

[11] CHULIN XIE ET AL., *PRIVACY OF AUTONOMOUS VEHICLES: RISKS, PROTECTION METHODS, AND FUTURE DIRECTIONS*, ABS/2209.04022 ARXIV (2022), HTTPS://API.SEMANTICSCHOLAR.ORG/CORPUSID:252185346.
[12] BLUEPRINT FOR AN AI BILL OF RIGHTS: MAKING AUTOMATED SYSTEMS WORK FOR THE AMERICAN PEOPLE, (2022).

Humans are likely to remain involved in some aspects of ADS operational safety and service. Two key human roles that may require policy attention are the *remote operator* and the *safety driver*. Humans in these supervisory roles can further or impede safety. If they fail to react to hazardous events resulting in *human-out-of-loop* issues that can lead to accidents or missteps. However, effective supervision and well-timed intervention can also yield safer operations, as observed in other high-risk complex engineering systems. However, automation that is too reliant on operators may pose challenges regarding human-system interaction safety.[13]

- **Safety drivers:** To protect the safety of driver workers and address safety risks, regulators could set guidelines or requirements for safety driving training and driver legal liability awareness. These include fatigue risk management policies or shift lengths, alerts to address automation complacency, and testing for automation complacency.
- **Remote Operations:** To protect remote operator safety and address broader safety risks, regulators could set guidelines or requirements for remote operation classifications, remote operator authority or control levels for activities, ensuring the functionality of remote operation stations, the format of information sent to the operators, set restrictions on the location of remote operators with respect to vehicles, and establish a risk evaluation rubric to assess remote operators/classification during fleet scaling.

Section IV. Conclusion

Key takeaways from this project are that ADS technologies are advancing rapidly. Meanwhile, engineers, regulators and legal scholars define safety differently and will likely need to better collaborate to advance the future of safe transportation systems. A blueprint for ADS Safety must be informed by experts in engineering, regulation, and law. This analysis arrived at two broad strategic ADS Safety approaches 1) reforming the Federal Motor Vehicle Safety Standards (FMVSS) or 2) the introduction of an ADS Safety Case Approach. The paper discusses options for ADS Data Policy, focusing on data collection and privacy considerations. The analysis concludes that data collection will be critical to inform oversight and guidance to ensure safe adoption and scaling of ADS, and to advance continuous improvement of ADS industry. Finally, the paper highlights the importance of considering human factors and human reliability in ADS safety, particularly in the roles of safety drivers and remote operators, where policy may aid in protecting human operators and address safety risks associated with human-system interaction. Overall, this white paper aims to provide stakeholders across sectors a foundational understanding of ADS safety and introduces policy directions for further exploration and development in this rapidly evolving field.

[13] MARILLIA A. RAMOS ET AL., *ACCOUNTING FOR HUMAN FAILURE IN AUTONOMOUS SHIP OPERATIONS*, *IN* SAFETY AND RELIABILITY – SAFE SOCIETIES IN A CHANGING WORLD 355 (2018), HTTPS://WWW.TAYLORFRANCIS.COM/CHAPTERS/OA-EDIT/10.1201/9781351174664-45/ACCOUNTING-HUMAN-FAILURE-AUTONOMOUS-SHIP-OPERATIONS-RAMOS-UTNE-VINNEM-MOSLEH.

Contents

I. Introduction

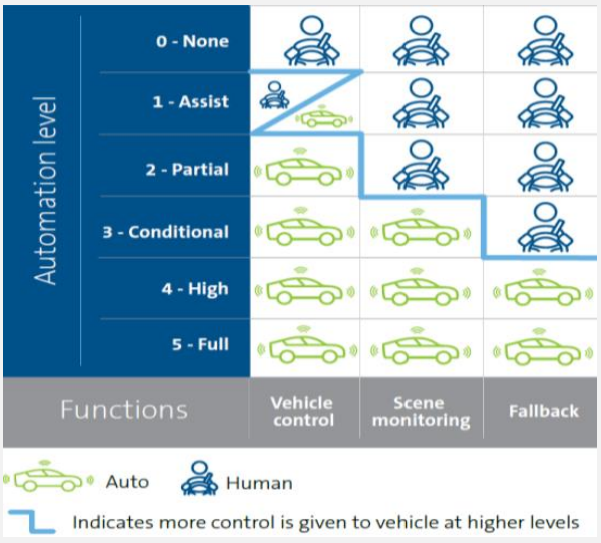
This white paper is as a primer for policymakers, industry representatives, and other policy stakeholders who seek to better understand automated driving system (ADS) safety, why safety regulations matter, what relevant safety regulations exist, and how ADS may herald significant changes in the safety regulatory landscape. This white paper focuses on highly automated vehicles, referred to here as ADSs, or ADS-equipped vehicles. The focus of this paper is not on automated driving assistance systems (ADAS) (See Figure 1). ADAS vehicles may require additional policy discussion, but these systems are out of scope of this white paper.

Organization of this Report

The paper is organized as follows: In the Introduction, authors define ADS at a high level. In the Background section (Section II) explores multidisciplinary definitions of *safety* through three lenses, 1) engineering standards setting organizations 2) regulatory and 3) legal scholars and the courts. Section III includes a summary of the U.S. federal administrative approach to regulating ADS safety and compares this approach to the California regulatory process (Additional international case studies are included in Appendix A.). The goal of Section II and III is to open silos for engineers, lawyers, and policy makers to view other approaches to safety definitions and investigate how norms differ across sectors. Section IV identifies key questions for ADS safety, data collection, and human-machine interfacing. The concluding Section V summarizes the analysis and suggests opportunities for further research.

Figure 1. Expanded SAE Levels of Automation.

Engineers' (SAE) six levels of driving automation range from no automation at all (SAE Level 0) to full automation (SAE Level 5). Criteria for differentiating between levels is shown at right. The definition of these levels have been widely adopted, but found to be lacking, especially in distinguishing among vehicles that are between levels. For example, vehicles with Automated Driving Assistance System (ADAS), also called partially automated vehicles, are referred to as Level 2 (L2) or L2+ vehicles.



What are Automated Driving Systems (ADSs)?

Automated Driving Systems can perform all or part of driving tasks on a sustained basis within specified operational conditions.[14] These systems include a set of elements that offer a specific conditional or higher automated driving use case for a specific *operations design domain* (*ODD).* ADSs vary in what software and hardware they use to detect and respond to events on the road. The conventional ADS applies a modular design pipeline, with three operational procedures occurring in a sequence: perception, prediction, and planning. [15]

In the conventional approach, first the perception phase uses inputs from a combination of sensors, such as cameras, radar, lidar, inertial sensors, and GPS. These sensors enable vehicles to carry out real-time perception of their environment, including categorizing other vehicles, pedestrians, or objects. This information will inform the prediction phase, which can determine the vehicles and other road users' next movement, and finally trajectory planning where the ADS estimates the outcomes of this movement.[16] Research efforts have also focused on developing an end-to-end perception to planning pipeline, aiming to achieve higher accuracy and adaptability of the ADS.

Different AI methods can be used to train ADS systems to perform these functions, including *imitation learning* and *reinforcement learning*. Imitation learning usually involves human experts (i.e., training, or expert data) to demonstrate how the system should behave in different scenarios using simulation software and on-road driving. [17] Other strategies include *Reinforcement Learning*, which can be complementary to imitation learning because it can use expert training data and other data sources to train the system to explore scenarios and determine what actions result in unsafe outcomes, like collisions, and what actions result in safe outcomes.[18] This training usually occurs offline, or in testing, and these strategies are not typically continuously employed while the vehicle is in operation, but new "knowledge" achieved in training is reviewed and tested before developers apply software updates.

Recent advances in this field include *Deep reinforcement learning*, which are models built through neural network architectures. When integrated into the ADS's planning tasks, these models make predictions and determine the vehicle's trajectory and planning actions.[19] [20] Deep reinforcement learning can improve accuracy

---

[14] 14:00-17:00, *ISO 34501:2022*, ISO (2022), HTTPS://WWW.ISO.ORG/STANDARD/78950.HTML.
[15] LI CHEN ET AL., *END-TO-END AUTONOMOUS DRIVING: CHALLENGES AND FRONTIERS*, (2024).
[16] EMMANUEL OWUSU APPIAH & SOLOMON MENSAH, *OBJECT DETECTION IN ADVERSE WEATHER CONDITION FOR AUTONOMOUS VEHICLES*, 83 MULTIMEDIA TOOLS AND APPLICATIONS 28235 (2024).
[17] JOONWOO AHN, MINSOO KIM & JAEHEUNG PARK, AUTONOMOUS DRIVING USING IMITATION LEARNING WITH LOOK AHEAD POINT FOR SEMI STRUCTURED ENVIRONMENTS, 12 SCIENTIFIC REPORTS 21285 (2022).
[18] YIREN LU ET AL., IMITATION IS NOT ENOUGH: ROBUSTIFYING IMITATION WITH REINFORCEMENT LEARNING FOR CHALLENGING DRIVING SCENARIOS, ARXIV PREPRINT ARXIV:2212.11419 (2022).
[19] ANDREAS FOLKERS, MATTHIAS RICK & CHRISTOF BÜSKENS, *CONTROLLING AN AUTONOMOUS VEHICLE WITH DEEP REINFORCEMENT LEARNING*, *IN* 2019 IEEE INTELLIGENT VEHICLES SYMPOSIUM (IV) 2025 (2019).
[20] SHENGBO EBEN LI, *DEEP REINFORCEMENT LEARNING*, *IN* REINFORCEMENT LEARNING FOR SEQUENTIAL DECISION AND OPTIMAL CONTROL 365 (2023), HTTPS://DOI.ORG/10.1007/978-981-19-7784-8_10.

for high complexity tasks,[21] but these systems are computationally intense and can be costly to acquire sufficient expert data to train and validate the models. A new technology goes even further, expanding neural network-based approaches using world foundation models, which operate by integrating large-language models (LLMs) into the ADS. World foundation models promise lower training periods and more accuracy.[22]

At the time, many different architectures for ADS-equipped vehicles are under development, testing, and early deployment stages. The choices of what AI tools to employ will differentiate players in the ADS market, as each looks to optimize accuracy and reduce costs. Each of these methods require significant computing power. It is within this context of constant and rapid technological change in AI and computing that regulators are tasked with setting guardrails for safety for AI and ADS systems. This paper aims to inform policy makers and practitioners.

[21] ID.
[22] HAOXIANG GAO ET AL., *A SURVEY FOR FOUNDATION MODELS IN AUTONOMOUS DRIVING*, (2024), HTTPS://ARXIV.ORG/HTML/2402.01105V1.

II. Safety Background

Safety can be considered more as a shared interpretation of individual or collective risk tolerance, rather than a static concept. Blumenthal et al. explain, "[t]he concept of safety is contextually, technologically, and culturally dependent."[23] Safety can be a measurement, safety can be a process, safety can also be a threshold comparing ADS safety performance to a human safety standard.[24]

Safety is all these things, and yet codified definitions for safety may reflect and project cultural norms for behavior. Paradoxically, safety is also said to be a culture in and of itself. Numerous safety experts claim an organization's safety management system requires a *safety culture*. Key elements of a *safety culture* include, "shared and safety relevant ways of thinking and acting," which can include "informal and formal organizational measures to achieve safety in organizations."[25]

Safety experts for ADS also expand on cultural expectations for safety by pointing to the concept of *roadsmanship*, which is defined as the ongoing act of safely driving on roads in such a way that a system or vehicle responds safely to hazards and does not create any hazards for others.[26] Shladover and Nowakowski focus on two additional dimensions of automotive safety, "functional safety with respect to internal faults"[27] and "driving behavior competency to deal with external hazards in the driving environment."[28] This view differentiates between inside and outside influences on safety.

Kurani differentiates between *safety* and *security*, which is especially useful when considering the roles of passengers, drivers, and potential cybersecurity criminals, or other bad actors. Kurani states that safety is "the condition of being secure from accidental harm" while "security is defined to be the condition of being safe from intentional harm."[29] Kurani is expanding on earlier work from Grundwald that addresses societal risk constellations of people interacting with AVs. Risk constellations are scattered perceptions of risk on different conceptual axes. For example, what constitutes a reasonable risk for a child will be different for an adult, and

[23] MARJORY S. BLUMENTHAL ET AL., *SAFE ENOUGH: APPROACHES TO ASSESSING ACCEPTABLE SAFETY FOR AUTOMATED VEHICLES*, (2020), HTTPS://WWW.RAND.ORG/PUBS/RESEARCH_REPORTS/RRA569-1.HTML
[24] MARJORY S. BLUMENTHAL ET AL., *SAFE ENOUGH: APPROACHES TO ASSESSING ACCEPTABLE SAFETY FOR AUTOMATED VEHICLES*, (2020), HTTPS://WWW.RAND.ORG/PUBS/RESEARCH_REPORTS/RRA569-1.HTML
[25] HAUKELID, *SUPRA* NOTE 7.
[26] *ID*.
[27] STEVEN E. SHLADOVER & CHRISTOPHER NOWAKOWSKI, *REGULATORY CHALLENGES FOR ROAD VEHICLE AUTOMATION: LESSONS FROM THE CALIFORNIA EXPERIENCE*, 122 TRANSPORTATION RESEARCH PART A: POLICY AND PRACTICE 125 (2019).
[28] *ID*.
[29] KENNETH S. KURANI, *USER PERCEPTIONS OF SAFETY AND SECURITY: A FRAMEWORK FOR A TRANSITION TO ELECTRIC-SHARED-AUTOMATED VEHICLES* (2019), HTTPS://ESCHOLARSHIP.ORG/UC/ITEM/40G1637B

risk constellations may vary based on age, gender, regional or cultural norms, etc.[30] With all this variance, safety must be thought of as a process and iterative exercise, rather than a static concept.

A. Safety Defined by Consensus Standards in Engineering

For engineers, aligning several consensus-based standards can make assessments of safety easier, and these standards intentionally provide flexibility to recognize that the world is full of unknowns and compromises between different stakeholders. Consensus-based standards setting organizations, like the Society of Automotive Engineers (SAE) and the International Organization for Standardization (ISO) form the backbone for safety assessments across many sectors, setting standards for automotive, aviation, health care, consumer goods, among many others. Standards organizations serve an important function, yet they are also criticized for favoring flexible standards (instead of strict standards) as their priority is to achieve consensus among diverse decision-makers. Flexible standards are preferred by industrial innovators to move innovations quickly, allowing them to iterate designs, without rendering standards obsolete. It can take several years between updates of standards. Regulatory agencies, whose standards-setting processes are often slower, often lean on the outcomes of standard-setting organizations to establish baseline protocols, especially where there is widespread consensus that a given approach is the best practice.

Safety and Risk

Standards setting organizations base their definitions of safety on long-standing definitions of risk. One of the most used definitions of risk among risk analysts is known as "the triplet of risk", formalized by Kaplan and Garrick as [31]

$$R = \{s_i, p_i, x_i >\}, \qquad i=1, 2, \ldots N.,$$

where s_i is a scenario identification or description;
p_i is the probability of that scenario; and
x_i, is the consequence or evaluation measure of that scenario, i.e., the measure of damage.

ISO, one of the foremost standard setting organizations, and considered a gold standard for definitions of safety, defines *risk* as the "effect of uncertainty on objectives,"[32] ISO 3100 applies a simple calculation,

Risk Index = Impact of Risk x Probability of Occurrence

[30] ARMIN GRUNWALD, SOCIETAL RISK CONSTELLATIONS FOR AUTONOMOUS DRIVING. ANALYSIS, HISTORICAL CONTEXT AND ASSESSMENT, IN AUTONOMOUS DRIVING: TECHNICAL, LEGAL AND SOCIAL ASPECTS 641 (MARKUS MAURER ET AL. EDS., 2016), HTTPS://DOI.ORG/10.1007/978-3-662-48847-8_30.
[31] Stanley Kaplan & John B. Garrick, *On The Quantitative Definition of Risk*, Vol. I, No. I, RISK ANALYSIS (1981), chrome-extension://efaidnbmnnnibpcajpcglclefindmkaj/https://www.nrc.gov/docs/ML1216/ML12167A133.pdf.
[32] INTERNATIONAL STANDARDS ORGANIZATION, *RISK MANAGEMENT — GUIDELINES*, (2018), HTTPS://WWW.ISO.ORG/OBP/UI/#ISO:STD:ISO:31000:ED-2:V1:EN.

ISO periodically updates safety definitions through a committee, or consensus-based, process. ISO currently defines safety in alignment with U.S. federal regulators as the "absence of unreasonable risk (AUR)."[33]

Based on these definitions of risk, risk analysis consists of answering: i) what can happen?, ii) how likely is that to happen?, and iii) if it does happen, what are the consequences? Safety risk analysis is a central exercise to inform how an ADS manufacturer can identify where it is most essential to invest time and resources, e.g., additional training and testing of their models. This allows ADS manufacturers to develop or interpret an infinite number of risk indices, setting specific risk management and optimization thresholds, based on what they determine to avoid "unreasonable risk" to their assets, shareholders, and the broader public. However, ultimately organizations have considerable latitude to interpret these flexible standards and create their own risk analysis strategies. Each ADS manufacturer or operator can claim compliance with the ISO standard and still assign residual risk slightly differently.

Based on these definitions, risk analysis and management can be considered as much a technical question as a business or ethical approach, so setting flexible metrics reflects an industry-wide willingness to allow different approaches to taking on risk, and different decisions about how to get risks lower.

Given these broad and flexible interpretations standards-organizations have also established boundaries for defining safety. SAE aligns their perception of safety with the ISO definition, and defines a related term, *safety of the intended functionality (SOTIF)* as the "absence of unreasonable risk due to hazards resulting from functional insufficiencies of the intended functionality or from reasonably foreseeable misuse by persons."[34] This definition raises additional concepts which recognize that there are limitless unknowns. SOTIF recognizes that a product is often considered safe for its intended function while operating as planned. A kitchen knife manufacturer may not be held fully responsible in the event a knife is used in a knife-throwing incident that results in injury. However, the risk that consumers misuse the product is expected to be reasonably prevented and/or mitigated through appropriate design, communication, or other strategies.

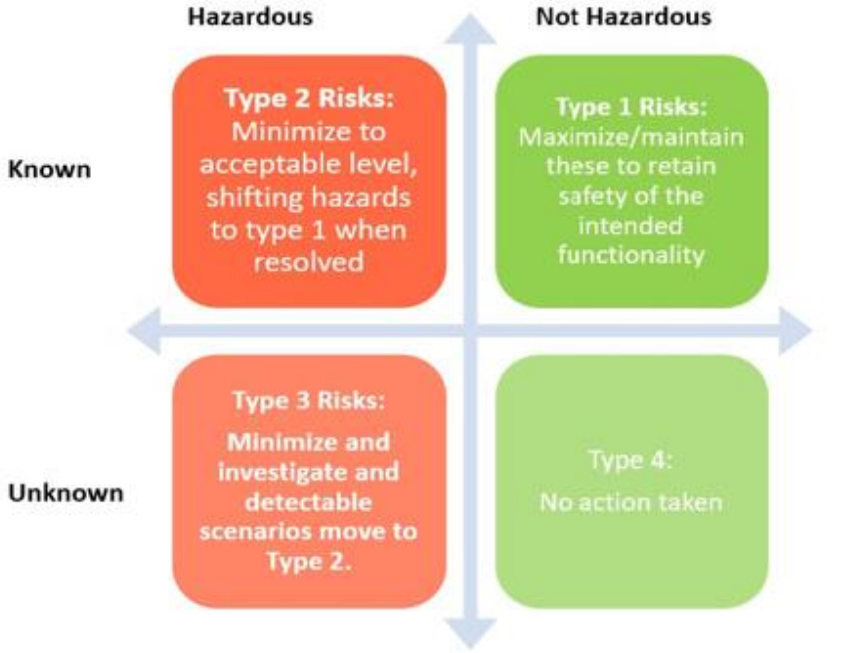
Warg et. al. recommend the use of a tailored quantitative risk norm (QRN) standard to assess ADS safety.[35] This would classify every conceivable incident type to a consequence class within the QRN, where each of

[33] ISO 26262-1:2018 ROAD VEHICLES — FUNCTIONAL SAFETY - PART 1: VOCABULARY, (2018), HTTPS://WWW.ISO.ORG/STANDARD/68383.HTML.
[34] CENTRE FOR CONNECTED & AUTONOMOUS VEHICLES BRITISH STANDARDS INSTITUTE, CONNECTED AND AUTOMATED MOBILITY – VOCABULARY, (2023), HTTPS://WWW.BSIGROUP.COM/GLOBALASSETS/LOCALFILES/EN-GB/CAV/PASS-AND-FLEX-PDFS/BSI-FLEX-1890-V5.0-FINAL-PDF_WATERMARK.PDF.
[35] FREDRIK WARG ET AL., THE QUANTITATIVE RISK NORM - A PROPOSED TAILORING OF HARA FOR ADS, IN 2020 50TH ANNUAL IEEE/IFIP INTERNATIONAL CONFERENCE ON DEPENDABLE SYSTEMS AND NETWORKS WORKSHOPS (DSN-W) 86 (2020).

alongside safety-related risks in the QRN framework, considering a broader range of incidents that may be relevant to local government regulators. Similar research from Ramos et al. includes a probabilistic risk assessment for highly automated passenger service, cataloging hazards using an event sequence diagram (ESD) and a fault tree analysis (FTA). [36] [37] [38]

Alexiades synthesizes the common standards and proposes a method to categorize risks into four quadrants. (see Figure 2). Type 1 and 2 risks are known and can be addressed and minimized where necessary, while type 3 and 4 risks are unknown.

Figure 2. Example of Risk Categorization Rubric



Source: Alexiades, N. (2021)

Safety Case Approach

For complex systems such as ADSs, publicly articulating all SOTIF, safety risks, and assumptions can provide transparency and accountability for consumers, regulators, and courts. Some consensus is building around the benefits of a risk-informed *safety case* approach for ADS. A safety case is simply "a structured argument supported by evidence, to justify that a system is acceptably safe for a specific application in a specific operating environment" [39] [40] A safety case is typically a document (sometimes hundreds or even thousands of pages) that lists a series of main safety claims, with supporting evidence, then divides these claims into subclaims with more detailed evidence, and then further subdivides each subclaim again until all known hazards can be shown to be recognized and mitigated (and limitations or unknowns explicitly stated).

[36] CAMILA CORREA-JULLIAN, M. RAMOS, A. MOSLEH, AND J. MA, "OPERATIONAL SAFETY HAZARD IDENTIFICATION METHODOLOGY FOR AUTOMATED DRIVING SYSTEMS FLEETS," PROC. INST. MECH. ENG. PART O J. RISK RELIAB., FEB. 2024, DOI: 10.1177/1748006X241233863

[37] CHRISTOPH THIEME ET AL., PROCEEDINGS OF THE 3RD INTERNATIONAL WORKSHOP ON AUTONOMOUS SYSTEMS SAFETY (2023).

[38] RAMOS, M.; THIEME, C.; UTNE, I.; MOSLEH, A. A GENERIC APPROACH TO ANALYSING FAILURES IN HUMAN – SYSTEM INTERACTION IN AUTONOMY. SAFETY SCIENCE, V. 129. SEPTEMBER 2020. HTTPS://DOI.ORG/10.1016/J.SSCI.2020.104808.

[39] BRITISH STANDARDS INSTITUTE, *SUPRA NOTE 31.*

[40] A-P-T RESEARCH, SAFETY CASE WORKSHOP (2014).

Safety Board.[41] Myklebust, an expert who evaluated and contributed to dozens of safety case documents in different sectors, explains that the purpose of developing a safety case early in technological development is to establish a "Minimum Viable Product Safety Case."[42]

Underwriter Laboratories (UL) has been developing safety standards since 1903 and has developed many international standards through a consensus process of a panel of experts (which is largely made up of industry representatives with some government and academic representatives). In 2019 UL published a draft safety case, *UL 4600: Standard for Safety for the Evaluation of Autonomous Products*.[43] The creators of UL 4600 developed several "prompts," which they summarize with the quip, "UL 4600: Did you do enough" and "#Didyouthinkofthat?" UL 4600 is a checklist. It does not specify *how* you might accomplish these safety goals, and this is intentional to provide flexibility in approach.[44]

UL 4600 prompts include:

- travel infrastructure (e.g., types of road surfaces, etc.),
- environmental effects (e.g., weather, or illumination, etc.),
- road obstructions (police traffic blocks, sand/salt trucks, street sweepers, etc.),
- specific road user rules, (e.g., motorcycles, oversize systems, emergency response systems), and
- vulnerable populations (pedestrians, scooters, other at-risk road users) [45]

Safety Standards such as UL 4600 and ISO 26262 can inform risk assessments and enable developers and engineers to categorize risks.[46]

B. Safety Defined by Regulators

Federal Approach to Safety

The definitions of *safety* and *safety standard*, according to U.S. federal statute (49 U.S.C. § 30102) are as follows:

> *"motor vehicle safety" means the performance of a motor vehicle or motor vehicle equipment in a way that protects the public against unreasonable risk of accidents occurring because of the*

[41] *ID.*
[42] THOR MYKLEBUST, FUTURE CHALLENGES, PITFALLS, AND OPPORTUNITIES WHEN USING A SAFETY CASE APPROACH FOR SW-INTENSIVE SYSTEMS, (2023). SLIDE 6
[43] PRESENTING THE STANDARD FOR SAFETY FOR THE EVALUATION OF AUTONOMOUS VEHICLES AND OTHER PRODUCTS | UL STANDARDS & ENGAGEMENT (ULSE.ORG)
[44] UNDERWRITERS LABORATORIES, UL 4600: GENERAL STAKEHOLDER OVERVIEW, (2019), HTTPS://COLLATERAL-LIBRARY-PRODUCTION.S3.AMAZONAWS.COM/UPLOADS/NFP/NFP_ASSET/ATTACHMENT/1424/191010_UL4600_POLICY_WEBINAR.PDF.
[45] ID.
[46] NICHOLAS ALEXIADES, *AV SAFETY STANDARDS FRAMEWORKS: HOW UL 4600, ISO 21448 (SOTIF) AND ISO 26262 CAN WORK TOGETHER TO BUILD A SAFETY CASE*, (2021), HTTPS://THE26262CLUB.COM/WP-CONTENT/UPLOADS/2021/04/5-NICHOLAS-ALEXIADES_UL_AV-SAFETY-STANDARDS-FRAMEWORKS_032421.PDF.

design, construction, or performance of a motor vehicle, and against unreasonable risk of death or injury in an accident, and includes nonoperational safety of a motor vehicle.[47]

All vehicle manufacturers, including ADS manufacturers, submit compliance with the relevant federal motor vehicle safety standards (FMVSS). Among the more than 70 FMVSS, this includes crashworthiness tests administered at independent testing sites, and some standards require manufacturers to attest to or demonstrate with photos how vehicles meet the standard. No permit is issued by USDOT for road safety compliance. Once an ADS or traditional original equipment manufacturer (OEM) certifies compliance with the FMVSS, with no known failures at testing sites, they can operate in the U.S. (if they additionally meet state vehicle compliance requirements or permits).[48] After this submission, regulators can intervene and issue recalls when a component fault (e.g. steering wheel, seatbelt, or software) is shown to be non-compliant or poses an "unreasonable risk" to public safety, which is defined as resulting in a "'non-negligible' number of crashes, injuries, or deaths…" [49] [50] Regulators can also open investigations for any reason and can test vehicles at random in the vehicle market to audit reported results.

NHTSA's main muscle is their reactive recall authority. When the agency receives a "non-negligible" number of complaints to indicate that a safety-related defect may be present in a vehicle, it may conduct an investigation to determine whether the vehicle poses an unreasonable safety risk or fails to meet a minimum safety standard.[51] Only after this determination can NHTSA impose a recall. The Transportation Recall Enhancement, Accountability, and Documentation ("TREAD") Act was passed in 2000 to require manufacturers to report information that may indicate a safety defect, however the language in that Act does not address software or cybersecurity issues.[52] NHTSA's self-certification and recall process may be unable to respond to quickly deployed over-the-air software updates. Legal scholar Himes points out that outdated requirements for notifying NHTSA by "first-class mail within a reasonable time is out of date" are insufficient, and she suggests that increased transparency for consumers, regulators, and manufacturers is necessary.[53] Himes suggests that NHTSA could institute a rulemaking around over-the-air software updates and investigate how standardizing these operating systems and components could improve compatibility and enable more responsive regulatory oversight.[54]

---

[47] MOTOR VEHICLE SAFETY, 49 USC § 30102(A)(9) (2018).
[48] NATIONAL HIGHWAY TRAFFIC SAFETY ADMINISTRATION, UNDERSTANDING NHTSA'S REGULATORY TOOLS INSTRUCTIONS, PRACTICAL GUIDANCE, AND ASSISTANCE FOR ENTITIES SEEKING TO EMPLOY NHTSA'S REGULATORY TOOLS, (2017), HTTPS://WWW.NHTSA.GOV/SITES/NHTSA.GOV/FILES/DOCUMENTS/UNDERSTANDING_NHTSAS_CURRENT_REGULATORY_TOOLS-TAG.PDF.
[49] *NHTSA ENFORCEMENT GUIDANCE BULLETIN 2016-02: SAFETY-RELATED DEFECTS AND AUTOMATED SAFETY TECHNOLOGIES*, 65705, HTTPS://WWW.FEDERALREGISTER.GOV/DOCUMENTS/2016/09/23/2016-23010/NHTSA-ENFORCEMENT-GUIDANCE-BULLETIN-2016-02-SAFETY-RELATED-DEFECTS-AND-AUTOMATED-SAFETY-TECHNOLOGIES.
[50] MOTOR VEHICLE SAFETY, 49 USC § 30102(A)(9).49
[51] SAFETY ISSUES & RECALLS, NAT'L HIGHWAY TRAFFIC SAFETY ADMIN., HTTPS://WWW.NHTSA.GOV/RECALLS (LAST VISITED AUG. 8, 2023)
[52] EMMA HIMES, *NHTSA UP IN THE CLOUDS: THE FORMAL RECALL PROCESS & OVER-THE-AIR SOFTWARE UPDATES*, 28 MICH. TECH. L. REV. 153, 155 (2021).
[53] ID. AT 167, 174.
[54] ID. AT 174.

Over-the-air updates are a broader issue than ADS, because ADAS vehicles, and other consumer electronics, may also use over-the-air updates. The issue as it relates to ADS is more specifically how USDOT can build capacity for software recalls. USDOT may need to build capacity to externally validate the effectiveness of a change to software, e.g., a change to the deep reinforcement model, or a neural network training that corrects for an error. This may help the public accept that they are effectively intervening to protect public safety.

However, the only way Federal regulators will know for certain if a software recall is effective is by waiting to see if federal accident data demonstrates that the issues is resolved, through no reported recurrence of the issue. NHTSA has collected empirical data on traffic accidents since the 1970's, using police-reported crash reporting data that captures the majority of serious incidents. Sampling techniques may also help to assess trends in crashes. These may, for example, estimate the incidence of unreported crashes as well as police-reported.[55] USDOT also began collecting accident data from multiple sources. Reports can come from ADS developers directly, police-reported accident data, or other sources. The USDOT uses this data to provide a public dashboard for reporting ADS crash data.[56]

California Regulatory Process

Over fifty ADS developers have tested or are operating out of California, although far fewer are deploying commercially. A permit is required for ADS testing or operation in California, and as of January 2024, only Mercedes-Benz, Nuro, and Waymo are permitted to deploy commercially within California, in a specified ODD.[57] A Half dozen companies are permitted for driverless testing of their ADSs, and several dozen have a permit to test with a safety driver.[58] Operators can apply to any of these permits in any order, but typically operators begin testing with safety drivers, then apply to test without drivers, then apply to advance to deployment.[59] These permits applications require evidence of compliance with FMVSS, or approved exemptions, as well as demonstrating safe operation in a narrowly specified ODD and must reapply for any permit to make changes to the ODD. The permitting process is a simple form, which requires ADS developers to report a series of self-attestations of road readiness and provide appendices as evidence of the claims. Permitted companies are also required to agree to provide data to the California DMV on how their fleet is operating, and when any accidents occur, and whether the ADS system was engaged at the time of the

[55] CRASH REPORT SAMPLING SYSTEM | NHTSA, HTTPS://WWW.NHTSA.GOV/CRASH-DATA-SYSTEMS/CRASH-REPORT-SAMPLING-SYSTEM (LAST VISITED MAY 13, 2024).
[56] STANDING GENERAL ORDER ON CRASH REPORTING | NHTSA, HTTPS://WWW.NHTSA.GOV/LAWS-REGULATIONS/STANDING-GENERAL-ORDER-CRASH-REPORTING (LAST VISITED JUN 8, 2023).
[57] AUTONOMOUS VEHICLE TESTING PERMIT HOLDERS, CALIFORNIA DMV, HTTPS://WWW.DMV.CA.GOV/PORTAL/VEHICLE-INDUSTRY-SERVICES/AUTONOMOUS-VEHICLES/AUTONOMOUS-VEHICLE-TESTING-PERMIT-HOLDERS/ (LAST VISITED JUN 9, 2023).
[58] AUTONOMOUS VEHICLE TESTING PERMIT HOLDERS, CALIFORNIA DMV, HTTPS://WWW.DMV.CA.GOV/PORTAL/VEHICLE-INDUSTRY-SERVICES/AUTONOMOUS-VEHICLES/AUTONOMOUS-VEHICLE-TESTING-PERMIT-HOLDERS/ (LAST VISITED JUN 9, 2023).
[59] CALIFORNIA AUTONOMOUS VEHICLE REGULATIONS, CALIFORNIA DMV, HTTPS://WWW.DMV.CA.GOV/PORTAL/VEHICLE-INDUSTRY-SERVICES/AUTONOMOUS-VEHICLES/CALIFORNIA-AUTONOMOUS-VEHICLE-REGULATIONS/ (LAST VISITED MAY 18, 2023).

incident.[60] ADSs in California that intend to operate a passenger ride service also must seek approval from the California Public Utilities Commission (CPUC).[61] The CPUC oversees passenger safety and ridehailing, while the state DMVs regulates safe vehicle operation, and NHTSA regulates vehicle safety.

Spotlight on the European Union and U.K.

In 2020 the United Nations Economic Commission for Europe (UNECE) adopted regulations governing automated lane keeping systems, mandating the system be "free of unreasonable risks," including not causing collisions that are reasonably foreseeable and preventable. The regulations cover transitions from automated system to human control, minimum risk maneuvers, data storage, and cyber security standards.[62] In June of 2023, the UN Economic Commission for Europe's Working Party 29 released revised guidelines for validating ADS.[63] The guidelines recognize that "[v]alidating ADS safety is a highly complex task which cannot be done comprehensively nor effectively through one validation methodology alone."[64] The guidelines propose a multi-pillar validation methodology composed of (a) a scenario catalog, (b) simulation/virtual testing, (c) track testing, (d) real world testing, (e) audit/assessment, and (f) in service monitoring and reporting. [65] Meanwhile, in 2022 the Centre for Connected and Autonomous Vehicles, under the U.K. Department for Transport, released a 2025 vision for connected and automated mobility technologies,[66] which outlines a safety framework to be expanded upon in the National Safety Principles. [67] For ADS authorization, vehicles must "[a]void collisions which a competent and careful driver could avoid" and "[t]reat other road users with reasonable consideration," while [a]void[ing] putting itself in a position where it would be the cause of a collision." [68] In-use regulation requires data to assess (1) "[w]hether the authorization remains valid," (2) "[h]ow frequently breaches are occurring, if at all," (3) "[w]hether sanctions are appropriate against the [authorized self-driving entity] for their vehicle operating below expectations," and (4) "[w]hether there has been a failure to disclose… safety-relevant information, which will be a criminal offense." [69]

[60] MOLLIE D'AGOSTINO ET AL., *CALIFORNIA AUTOMATED VEHICLE POLICY STRATEGIES* (2021), HTTPS://ESCHOLARSHIP.ORG/UC/ITEM/6S59C5B7 (LAST VISITED APR 16, 2023).
[61] CALIFORNIA PUBLIC UTILITY COMMISSION, *AUTONOMOUS VEHICLE PROGRAMS*, HTTPS://WWW.CPUC.CA.GOV/REGULATORY-SERVICES/LICENSING/TRANSPORTATION-LICENSING-AND-ANALYSIS-BRANCH/AUTONOMOUS-VEHICLE-PROGRAMS (LAST VISITED MAY 13, 2024).
[62] ECE/TRANS/WP.29/2020/81, HTTPS://UNDOCS.ORG/HOME/MOBILE?FINALSYMBOL=ECE%2FTRANS%2FWP.29%2F2020%2F81&LANGUAGE=E&DEVICETYPE=DESKTOP&LANGREQUESTED=FALSE
[63] (WP.29/GRVA) WORKING PARTY ON AUTOMATED/AUTONOMOUS AND CONNECTED VEHICLES (16TH SESSION) | UNECE, 29, HTTPS://UNECE.ORG/INFO/EVENTS/EVENT/374929.
[64] ID. AT 2.
[65] ID.
[66] GRANT SHAPPS MP SECRETARY OF STATE FOR TRANSPORT & KWASI KWARTENG MP SECRETARY OF STATE FOR BUSINESS, ENERGY AND INDUSTRIAL STRATEGY, *CONNECTED & AUTOMATED MOBILITY 2025: REALISING THE BENEFITS OF SELF-DRIVING VEHICLES IN THE UK*, (2022), HTTPS://ASSETS.PUBLISHING.SERVICE.GOV.UK/GOVERNMENT/UPLOADS/SYSTEM/UPLOADS/ATTACHMENT_DATA/FILE/1099173/CAM-2025-REALISING-BENEFITS-SELF-DRIVING-VEHICLES.PDF .
[67] ID. AT 32-34.
[68] ID. AT 37.
[69] ID.

ADS Regulatory Investigations

There are several examples where California or Federal regulators responded to ADS safety incidents. There is some duplication of investigatory safety responsibilities from California and federal authorities, which serve slightly different statutory mandates, but this duplication may lead to confusion for the public. One example includes Pony.ai, an ADS developer who had acquired a California permit to test without a driver. In October 2021 their test vehicle collided with a road center divider and a traffic sign. After the accident was reported, California DMV swiftly revoked their permit.[70] After this occurred NHTSA opened an investigation into Pony.AI's ADS,[71] finding that the vehicle crashed into a street sign within 2.5 seconds of the ADS operation shutting down due to, a confluence of rare errors involving, "small floating point number rounding errors (or "discrepancies"). As a result, it was possible that one specific diagnostic matching function could incorrectly interpret an inconsequential rounding discrepancy as a geolocation mismatch."[72] This series of errors caused the vehicle ADS system to turn off and it quickly collided with barriers in the roadway. In their recall safety report NHTSA recognized that Pony.ai had immediately addressed the erroneous code after the crash. For this reason, NHTSA stated that the 3 vehicles effected by the recall did not require permanent removal from the road and that the remedy that Pony.ai took was sufficient.[73]

However, given the state DMV permit revocation, these vehicles cannot operate in California until a new state permit is issued. As of October 2023, Pony.ai does not have a permit to operate driverless testing in California, however during the reapplication process they were denied a permit due to a failure to report accurate information on their safety drivers.[74] The company has since been reported to have been operating with a safety driver, in Tuscan, Arizona in 2022,[75] and has launched fully autonomously in several Chinese cities including Beijing and Guangzhou.[76]

In 2023, Cruise Automation was involved in an injury-involved collision initiated by a nighttime hit-and-run driver in downtown San Francisco which resulted in the Cruise vehicle colliding with the injured party. This accident prompted several regulatory inquiries and outcomes. The accident was reported to both California DMV and NHTSA, and Cruise briefed both agencies. Also in October, NHTSA opened an investigation, and both

---

[70] REBECCA BELLAN, *PONY.AI LOSES PERMIT TO TEST AVS WITH DRIVER IN CALIFORNIA*, TECHCRUNCH (MAY 24, 2022), HTTPS://TECHCRUNCH.COM/2022/05/24/PONY-AI-LOSES-PERMIT-TO-TEST-AUTONOMOUS-VEHICLES-WITH-DRIVER-IN-CALIFORNIA/ (LAST VISITED JUL 26, 2023).

[71] JOSEPH OXENHAM, *ADS DESIGNED AND MANUFACTURED BY PONY.AI*, (2022), HTTPS://STATIC.NHTSA.GOV/ODI/INV/2022/INOA-AQ22001-9345.PDF.

[72] *ID.*

[73] *ID.*

[74] MEGAN MYSCOFSKI, *SELF-DRIVING CAR STARTUP THAT LOST PERMIT IN CALIFORNIA COMING TO TUCSON*, ARIZONA PUBLIC MEDIA - NPR/PBS, SEP. 23, 2022, HTTPS://NEWS.AZPM.ORG/S/95269-SELF-DRIVING-CAR-STARTUP-THAT-LOST-PERMIT-IN-CALIFORNIA-COMING-TO-TUCSON/.

[75] ADAM STEINBERG & LAURA SHAW, *AUTONOMOUS DRIVING COMPANY PONY.AI EXPANDS INTO TUCSON, ARIZONA*, SUN CORRIDOR, HTTPS://SUNCORRIDORINC.COM/2022/09/22/AUTONOMOUS-DRIVING-COMPANY-PONY-AI-EXPANDS-INTO-TUCSON-ARIZONA/.

[76] PONY.AI TO LAUNCH FLEET OF ROBOTAXIS WITH ONTIME, GAC'S RIDE-HAILING APP, IN GUANGZHOU IN 2022, BUSINESSWIRE, APR. 26, 2022, HTTPS://WWW.BUSINESSWIRE.COM/NEWS/HOME/20220425006102/EN/PONY.AI-TO-LAUNCH-FLEET-OF-ROBOTAXIS-WITH-ONTIME-GAC%E2%80%99S-RIDE-HAILING-APP-IN-GUANGZHOU-IN-2022.

agencies received a briefing on the incident. California DMV ultimately acted first, revoking the Cruise permit to operate without a safety driver due to the safety risks raised, and due to claims that the company misrepresented or omitted pertinent information.[77] NHTSA issued a recall soon after along similar grounds. NHTSA claimed that Cruise's *Collision Detection Subsystem* should be recalled given, "the Cruise ADS inaccurately characterized the collision as a lateral collision and commanded the AV to attempt to pull over out of traffic, pulling the individual forward, rather than remaining stationary."[78] In this recall, Cruise attests that an update to the software involved is sufficient remedy, and a future replication of this specific circumstance, would result in a stationary vehicle response.[79]

These examples raise many questions. Were these federal and state regulatory interventions effective? Or were they duplicative? How might they be improved and scaled? How can the findings from these investigations result in industry-wide improvement?

C. Defining Safety: Legal Sector

While regulators and engineering standards organizations may be at the forefront of advancing industry safety standards, it is ultimately judges who will define how legal liability standards will apply to specific incidents.[80] Judges will also lay the foundation for civil and criminal sentencing for ADS involved legal proceedings. The U.S. judicial system is designed to apply statutory guidance to real circumstances, adding specificity and determining what constitutes reasonable safe operation of products and goods. U.S. courts will inevitably make decisions that will have significant impacts on liability, insurance, and thus market adoption and scaling of ADS technologies.

Liability in Robotics and AI

Legal scholars worldwide are looking into robot tort, as it applies to AI for use in drones, robotics in manufacturing, in surgical settings, and in numerous other consumer electronics products.[81] [82] There are different characteristics that make liability more difficult to assess for systems with AI elements, linked to the opacity of deep learning algorithms and unpredictability inherent to systems of deep learning which makes assessing causality more challenging.[83] Opacity, here is meant to capture several issues around inscrutability and lack of explainability of the decision-making process of deep learning models. In these models' developers

[77] CALIFORNIA DEPARTMENT OF MOTOR VEHICLES, *DMV STATEMENT ON CRUISE LLC SUSPENSION*, (2023), HTTPS://WWW.DMV.CA.GOV/PORTAL/NEWS-AND-MEDIA/DMV-STATEMENT-ON-CRUISE-LLC-SUSPENSION/.
[78] LLC CRUISE, *PART 573 SAFETY RECALL REPORT*, (2023), HTTPS://STATIC.NHTSA.GOV/ODI/RCL/2023/RCLRPT-23E086-7725.PDF.
[79] ID.
[80] JEFFREY J. RACHLINSKI & ANDREW J. WISTRICH, *JUDGING AUTONOMOUS VEHICLES*, 24 YALE JOURNAL OF LAW & TECHNOLOGY 707 (2022).
[81] ALICE GUERRA, FRANCESCO PARISI & DANIEL PI, *LIABILITY FOR ROBOTS I: LEGAL CHALLENGES*, 18 JOURNAL OF INSTITUTIONAL ECONOMICS 331 (2022).
[82] BRUYNE, GOOL, AND GILS, *SUPRA* NOTE 5.
[83] LLORCA ET AL., *SUPRA* NOTE 4.

can input training data but they may not be able to explain how a model interprets the data, or uses a learning algorithm, or which features are the most relevant.

However, even where there is opacity, our justice system relies on fault-based liability frameworks that enable judges to assign levels of fault to designers, manufacturers, deployers, and others responsible for negligence in maintaining, updating, or monitoring technology.[84] There have been some interesting cases, including a tragic case of a street-cleaning robot colliding with a baby carriage. In this case, Llorca et al. suggests that determining fault was complex. In addition to investigating whether flaws in the street-cleaning robot's perceptive abilities occurred (which could be linked to component error or assemblage error) there may have been adversarial cybersecurity issues (e.g. sensors were jammed or spoofed), or the human remote operator supervising could also carry some fault which could be systemic or individual in nature.[85]

In a similar evaluation Kropka finds that when robotics are used in medical surgeries, accidents have occurred, for example liability may be shared by a surgeon, a hospital, and a robotics manufacturer. Kropka arrives at an analogous conclusion for framing liability in automated vehicles (AVs) stating, "liability might extend to the manufacturer where they fail to adequately warn of the potential dangers of AVs, or where they overstate the capabilities."[86]

Other legal scholars have also evaluated the evolving role of liability in AI and robotics sectors. Guerra et al point out that the robots present new challenges for liability determinations and may need to be held to new standards:

> *Failing to account for every special circumstance cannot be regarded as a design flaw. However, we could design rules that might keep incentives in place for manufacturers to narrow the range of design limitations through greater investments in R&D and/or safety updates."* [87]

Judges will have latitude to assess not only that a defendant caused harm, or the extent to which that harm impacted the victim, but that the harm was caused by negligence, and non-compliance with the duty of care. In order to make this assessment, judges will need to evaluate technical documentation before and after an accident. This evidence may require external validation to testify to the veracity of the information, which will be a challenge given the nature of bespoke and proprietary software. And further, for any independent evaluator to demonstrate causal inference based on available information will also not be straightforward. [88]

---

[84] ID.
[85] ID.
[86] CAROLINE KROPKA, *"CRUISE"ING FOR "WAYMO" LAWSUITS: LIABILITY IN AUTONOMOUS VEHICLE CRASHES*, UNIVERSITY OF RICHMOND SCHOOL OF LAW: JOURNAL OF LAW & TECHNOLOGY (NOV., 2023), HTTPS://JOLT.RICHMOND.EDU/2023/11/23/CRUISEING-FOR-WAYMO-LAWSUITS/.
[87] GUERRA, PARISI, AND PI, *SUPRA* NOTE 77.
[88] LLORCA ET AL., *SUPRA* NOTE 6.

Considerations for Determining Liability in ADS Accidents

To successfully bring a product liability claim under U.S. state tort law, a plaintiff must show as an essential element that a "product was in a defective condition unreasonably dangerous to the user."[89] However, states are split on how to determine whether a product is "unreasonably dangerous." The majority of states use some form of the *risk-utility test*, which weighs the danger of the vehicle against its usefulness,[90] while the rest use a *consumer-expectations test*, which determines whether a "product did not perform as safely as an ordinary consumer would expect when used in the intended or reasonably foreseeable manner."[91] California uses a hybrid-approach, where risk-utility is applied when "the issue of design defect goes beyond the common experience of the product's users."[92] These standards can be difficult to apply when new technologies are involved. For example, the usefulness of a product may be difficult to assess if the market for that product is too small when considering the risk-utility test. Additionally, a consumer may not know what to expect from a newly developed technology which would sway consumer-expectation tests. Aggrieved consumers are not likely willing to excuse an ADS developer's misdeeds because they were in compliance with UL4600 and ISO 26262, instead the courts will focus will be on how they were harmed despite these actions.

More legal clarity on how courts will assess ADS safety will also be contingent on whether manufacturers arbitrate out of court or offer settlements to injured plaintiffs. For example, in the case of Oscar Nilsson, who crashed his motorcycle into General Motors' (GM) self-driving Cruise Chevy Bolt 2017 in San Francisco.[93] Nilsson was attributed total fault for the accident by the police report, yet he brought a successful civil lawsuit against GM and settled for undisclosed terms.[94] It is possible that this suit would have never been brought against GM given that the police officer writing the accident report assessed the situation and determined the plaintiff to be the negligent party. However, the new technology under discussion provided a window for the plaintiff to make a product liability dispute and disavow personal responsibility.[95]

The decision to resolve the Nilsson motorcyclist settlement may be linked to empirical biases, observed in how judges view ADS. In a study of 967 judges, researchers Rachlinski and Wistrich found "judges are biased against self-driving vehicles."[96] The study provided judges several scenarios and asked to award liability and treatment and award compensation, and they found that more liability was assigned to the self-driving vehicle than to a human in the same circumstances, and that higher damages were awarded to those plaintiffs involved in a

[89] *MADDEN V. COX*, 284 S.C. 574, 579 (1985).
[90] *BRANHAM V. FORD MOTOR CO.*, 390 S.C. 203, 220 NN.11, 218–219 (2010).
[91] *AUBIN V. UNION CARBIDE CORP.*, 177 SO. 3D 489, 504 (2015).
[92] *SOULE V. GENERAL MOTORS CORP.*, 8 CAL 4TH 548, 617 (1994).
[93] *MOTORCYCLIST SUES GM OVER CRASH WITH SELF-DRIVING CHEVY BOLT*, JALOPNIK (2018), HTTPS://JALOPNIK.COM/MOTORCYCLIST-SUES-GM-OVER-CRASH-WITH-SELF-DRIVING-CHEVY-1822358606.
[94] *SEE NILSSON V. GENERAL MOTORS LLC*, 4:18-CV-00471, (N.D. CAL. MAY 30, 2018) ECF NO. 32
[95] GARY MARCHANT & RIDA BAZZI, *AUTONOMOUS VEHICLES AND LIABILITY: WHAT WILL JURIES DO?*, 26 BOSTON UNIV. J. SCI. TECHNOL. LAW 67, 70 (2020).
[96] RACHLINSKI AND WISTRICH, *SUPRA* NOTE 76.

hypothetical accident involving an ADS. These observed biases reflect a human inclination to view novel technology in a negative or distrusting light.[97]

Considerations Regarding Duty of Care

AI technologies, and ADS in particular, will also test the legal concept of *duty of care*. *Duty of care* is a legal concept that assumes individuals are widely held to ethical norms of conduct. This concept is quite similar to the *roadsmanship* concept introduced in the previous section (page 11). D'Amato *et al.* offers a legal basis for how to interpret *duty of care* for use in cases involving an ADS, arguing ADSs much like all road users, are expected to abide by a social contract to take reasonable and prudent actions, even when doing so would violate traffic code, if doing so would avoid catastrophic outcomes.[98] [99] This means that there is a legally sound rationale (or even a tacit requirement) to break the law if doing so will prevent harm, yet this presents a conundrum that may play out differently in different circumstances.

Considerations for Criminal Cases involving ADS

Governments will also bring criminal suits against parties involved in ADS operation. For example, in the 2017 Tempe accident, safety driver Rafaela Vasquez was originally charged with negligent homicide when the Uber Advanced Technologies Group (ATG) ADS she was operating resulted in a fatal crash killing Elaine Herzberg. Vasquez pled guilty to a reduced crime, endangerment, and received a sentence of 3-years' probation. The victim's family and Uber ATG also settled an undisclosed civil suit privately out of court.[100] While there are technical differences between an ADS vehicle in testing mode with a back-up safety driver, and an ADAS vehicle, both legally would find that the driver is the responsible party if they are behind a steering wheel in the event of an accident.

The Valasquez incident was followed by a National Traffic Safety Board (NTSB) investigation that chastised NHTSA for a lack of action on AV safety testing. Further, the investigation found that Uber ATG's AV operations lacked a formal safety plan. While they used the term, *safety plan*, there are a range of definitions of what such a report would entail, given the lack of regulatory compliance requirements. NTSB found that, before the accident, Uber did not have a *fatigue risk management policy* for their safety drivers or personnel assigned to assess risks for safety drivers. NTSB's investigation also pointed to a human-machine interaction challenge

[97] ID.
[98] A D'AMATO ET AL., *EXCEPTIONAL DRIVING PRINCIPLES FOR AUTONOMOUS VEHICLES*, MOB. 2 UNIV. MICH. J. LAW MOBIL., HTTPS://FUTURIST.LAW.UMICH.EDU/EXCEPTIONAL-DRIVING-PRINCIPLES-FOR-AUTONOMOUS-VEHICLES/.
[99] J. CHRISTIAN GERDES & SARAH M. THORNTON, *IMPLEMENTABLE ETHICS FOR AUTONOMOUS VEHICLES, IN AUTONOMOUS DRIVING: TECHNICAL*, LEGAL AND SOCIAL ASPECTS 87 (MARKUS MAURER ET AL. EDS., 2016), HTTPS://DOI.ORG/10.1007/978-3-662-48847-8_5.
[100] ANDREW J. HAWKINS, *UBER DRIVER IN FIRST-EVER DEADLY SELF-DRIVING CRASH PLEADS GUILTY*, THE VERGE (AUG. 1, 2023), HTTPS://WWW.THEVERGE.COM/2023/7/31/23814474/UBER-SELF-DRIVING-FATAL-CRASH-SAFETY-DRIVER-PLEAD-GUILTY (LAST VISITED JAN 30, 2024)

referred to as *automation complacency*, which is an observed phenomenon that can occur after safety operators build trust with an ADS or ADAS vehicle's good performance and fail to monitor it effectively.[101]

Uber ATG took several actions after the NTSB report was released. They established a Safety and Responsibility Advisory Board to independently monitor safety throughout the organization. This Board included a former administrator at NHTSA, a former Boeing test pilot, as well as other safety experts. In 2020, they released a Safety Report which loosely applied a safety case framework (modeled after UL 4600 and other engineering standards).[102] They also established what they called a safety framework, which was a more comprehensive and forward-looking document, listing possible protocols such as "fatigue management strategies, limitations on use of electronic devices." Despite these efforts and a pivot towards improving *safety culture* and accountability, Uber ATG shuttered and was sold to Aurora later in 2020, and Aurora is currently marketing autonomous freight vehicles.[103]

Legal Considerations Regarding Privacy

Two additional areas of developing law deserve attention. Fourth Amendment protections, and their effects on individual privacy, and trade secrets, which consider corporate privacy. Both may limit what data regulators can collect when assessing ADS safety, but there are precedents for what data is collected and how it is shared. For corporate privacy, in some cases developing non-disclosure systems, like privately held data repositories may address issues, but there are types of data where non-disclosure would not be a potential remedy.

Personal Privacy Issues

For individuals, there are pervasive concerns about the sharing of too much mobility data, which can lead to reidentification, public exposure of sensitive information, targeted cyber-criminal activity, or the most severe outcome is stalking, or violence perpetrated on individuals due to data exposure. These concerns should be balanced against the need to collect information to ensure safety.

The Fourth Amendment provides protections against government collection of private information, but how these protections will be extended, if at all, to ADS systems will be the source of litigation. In considering regulation governing AV safety, it is important to consider the possibility that certain information may be barred from disclosure to a government agency under the Fourth Amendment. The following cases show the evolution of Fourth Amendment protections and can provide guidance:

[101] NATIONAL TRANSPORTATION SAFETY BOARD, COLLISION BETWEEN VEHICLE CONTROLLED BY DEVELOPMENTAL AUTOMATED DRIVING SYSTEM AND PEDESTRIAN (2018), HTTPS://DATA.NTSB.GOV/DOCKET/DOCUMENT/DOCBLOB?ID=40479021&FILEEXTENSION=.PDF&FILENAME=NTSB%20-%20ADOPTED%20BOARD%20REPORT%20HAR-19%2F03-MASTER.PDF.

[102] UBER ATG, *UBER ATG RELEASES 2020 SAFETY REPORT*, MEDIUM (AUG. 2020), HTTPS://MEDIUM.COM/@UBERATG/UBER-ATG-RELEASES-2020-SAFETY-REPORT-575DB33F2BD7.

[103] UBER, AURORA IS ACQUIRING UBER'S SELF-DRIVING UNIT, ADVANCED TECHNOLOGIES GROUP, ACCELERATING DEVELOPMENT OF THE AURORA DRIVER, UBER INVESTOR (DEC. 7, 2020).

- *Katz v. United States* has long been the seminal case on Fourth Amendment protections against government surveillance.[104] The Supreme Court established, and has since applied, that such protections exist where a person has a "reasonable expectation of privacy", and that exception is "one that society is prepared to recognize as 'reasonable'."[105]

- *United States v. Miller*[106] and *Smith v. Maryland* [107] developed the "third-party doctrine" exception to these Fourth Amendment protections, holding that there is no expectation of privacy in information voluntarily turned over to a third party. Since 2000, this exception has proven problematic to apply. The amount of information the average person now discloses to third parties was inconceivable at the time of these decisions in the 1960s and 1970s.

- *Carpenter v. United States* held special protections for cell site location information (CSLI) tracking. The Court analogized cell phones to "almost a 'feature of human anatomy'" and ruled that CSLI was not voluntarily shared information since it is a by-product of a cell phone linking to a cell tower.[108] Because of these factors, individuals would have a reasonable expectation of privacy and the data would not fall under the third-party doctrine. The Court specifically distinguished vehicles which "individuals regularly leave," but acknowledges the protections against long term GPS tracking established by *United States v. Jones*.[109] [110] The narrow holding in *Carpenter* was only applied to CSLI and with a 5-4 decision there is room for the current Supreme Court to rule differently on Fourth Amendment protections for AVs and how it applies to GPS data collected by AV companies.

- In *Sanchez v. L.A. Dep't of Transp.* the 9th circuit chose not to apply the *Carpenter* protections to e-scooter MDS location data collected by the Los Angeles Department of Transportation.[111] They held that in contrast to *Carpenter*, Sanchez voluntarily disclosed his location data because it is "a central feature of his transaction with [the] third party."[112] Also, the MDS data only tracks a user for the duration of their trip, unlike cell phone data tracking which is "virtually continuous."[113] Last, the court posits that "e-scooters, unlike cell-phones, are simply not 'indispensable to participation in modern society.'" [114]

[104] *KATZ V. UNITED STATES*, 389 U.S. 347 (1967).
[105] ID. AT 360–61.
[106] *UNITED STATES V. MILLER*, 425 U.S. 435 (1976).
[107] *SMITH V. MD.*, 442 U.S. 735 (1979).
[108] *CARPENTER V. UNITED STATES*, 138 S. CT. 2206, 2218 (2018).
[109] ID.
[110] *UNITED STATES V. JONES*, 565 U.S. 400 (2012). (THE PLANTING OF A GPS DEVICE ON A CAR IN A PUBLIC SPACE AND TRACKING ITS MOVEMENTS FOR 28 DAYS CONSTITUTED A "SEARCH" BASED ON THE COMMON-LAW TRESPASSORY TEST, HOWEVER THIS CASE DID NOT CONSIDER THE REASONABLE EXPECTATION OF PRIVACY TEST OR HOW IT MAY APPLY TO THE FACTS AT ISSUE).
[111] *SANCHEZ V. L.A. DEP'T OF TRANSP.*, 39 F.4TH 548 (2022).
[112] ID. AT 731.
[113] ID. AT 731–32.
[114] ID. AT 732. (QUOTING *CARPENTER V. UNITED STATES*, 138 S. CT. AT 2220)

Likely an analysis of *Carpenter* as it applies to data disclosed by transportation networking companies (ride-hailing) to government regulators would be similar to *Sanchez*, but it may be a closer call for personal, consumer AVs. The central questions to ask considering the 9th circuit's holding are (1) will AVs be "indispensable to participation in modern society," [115] (2) does the data collection constitute "pervasive track[ing] over an extended period," [116] and (3) is the data voluntarily disclosed by the user?[117] The answers to these questions will help inform whether *Carpenter* protections apply and whether potential AV consumer expectations of privacy are reasonable in the view of society.

Legal Considerations Regarding Trade Secrets

Legal scholars Tait and Katyal explain that trade secret claims may be used to conceal information that may have otherwise been disclosed.[118] Companies are establishing broader boundaries of trade secrecy, expanding it to cover "nontraditional subject matter" with only attenuated connection to competitive advantage.[119] Unlike copyrights or patents, which must first be registered with the government, trade secrets have a "self-defining character," where companies may assert confidentiality without pushback until the claim is adjudicated.[120] This allows companies to craft these claims to pursue the particular goal at hand.

Under California Civil Code § 3426.1, a trade-secret is,

> *…information, including a formula, pattern, compilation, program, device, method, technique, or process, that: (1) derives independent economic value, actual or potential, from not being generally known to the public or to other persons who can obtain economic value from its disclosure or use; and (2) is the subject of efforts that are reasonable under the circumstances to maintain its secrecy.*

Additionally, the Supreme Court has expanded the scope of the trade secret exemption from Freedom of Information Act requests. In the 2019 case *Food Marketing Institute v. Argus Leader Media*, the Court disposed of the prior test for confidentiality which required the disclosure to be "likely …to cause substantial harm."[121] Instead the Court allowed "[a]t least where commercial or financial information is both customarily and actually treated as private by its owner and provided to the government under an assurance of privacy," it may be considered confidential under the trade secrets exemption to Freedom of Information Act.[122] This may indicate a growing agreement by the courts with corporations' broader drawings of the boundaries around trade secrecy.

In 2021, Waymo applied for a permit from the DMV to deploy their AVs on public streets in California. In this application, Waymo was required to disclose to the DMV information relating to their AV technology, including

---

[115] *CARPENTER V. UNITED STATES*, SUPRA NOTE 104 AT 2220.
[116] *SANCHEZ V. L.A. DEP'T OF TRANSP.*, SUPRA NOTE 107 AT 560.
[117] ID. AT 559.
[118] SONIA KATYAL & CHARLES GRAVES, *FROM TRADE SECRECY TO SECLUSION*, SSRN JOURNAL (2021), HTTPS://WWW.SSRN.COM/ABSTRACT=3760123.
[119] ID. AT 1342.
[120] ID. AT 1350–51.
[121] *FOOD MKTG. INST. V. ARGUS LEADER MEDIA*, 139 S. CT. 2356, 2360–61 (2019).
[122] ID. AT 2366.

"operational processes and design capabilities.[123] Later that year, a California Public Records Act request was made by a suspected news organization requesting this information, and an unredacted version was shared. Waymo successfully argued in the California Superior Court this information constituted a trade secret and "[i]nformation related to the safety processes integrated into Waymo's AV design and operation... would allow competitors to adopt for themselves those processes, and to glean other operational intelligence by analyzing the criteria Waymo uses to assume a low-risk operating condition."[124] The court barred disclosure under the California Public Records Act of all the information that Waymo claimed as trade secrets for two years to any third party.

California is deliberating who constitutes a third party. There is a history of disputing data reporting requirements due to concerns about trade secrets.[125] Confidential declarations for the California Public Utilities Commission (CPUC) quarterly data reports show that trade secrets are already a basis for redacting important information to inform AV safety policy, but the CPUC has waffled on whether to allow trade secrets. A CPUC administrative Law Judge made a decision rejecting some claims for data redaction, and then later recalled this decision. Trade secret claims by Lyft and Uber, are ongoing, and the companies are pursuing legal avenues to redact trip data, which would keep this information confidentially reported. However, the ALJ also rejected data redaction requests on other grounds, including privacy. [126]

This reversal speaks to the challenges that these regulators have faced while navigating complex privacy issues. Many regulators lack expertise in data confidentiality techniques, and have limited experience in solving data privacy issues. Future research can strive to provide clarity that helps decision-makers better assess the tradeoffs when considering competing calls for data confidentiality and transparency.

[123] OPENING BRIEF FOR PLAINTIFF, *WAYMO LLC, V. CAL. DMV*, NO. 34-2022-80003805-CU-WM-GDS (CAL. SUPER. CT. FILED FEB. 2, 2022), 7.
[124] ID.
[125] *DECISION REQUIRING TRANSPORTATION NETWORK COMPANIES TO SUBMIT THEIR ANNUAL REPORTS FOR THE YEARS 2014-2019 TO THE COMMISSION WITH LIMITED REDACTION*, R.12-12-001 (CAL. P.U.C. SEPT. 30, 2022) HTTPS://DOCS.CPUC.CA.GOV/SEARCHRES.ASPX?DOCFORMAT=ALL&DOCID=497334003
[126] CALIFORNIA PUBLIC UTILITY COMMISSION, *QUARTERLY REPORTING*, HTTPS://WWW.CPUC.CA.GOV/REGULATORY-SERVICES/LICENSING/TRANSPORTATION-LICENSING-AND-ANALYSIS-BRANCH/AUTONOMOUS-VEHICLE-PROGRAMS/QUARTERLY-REPORTING (LAST VISITED MAY 13, 2024).

III. ADS Safety Policy Considerations

In this section, we expand on several topics raised in the *Blueprint for an AI Bill of Rights* issued by the White House in 2022, that aptly frame the ADS safety discussion (shown in Table 1.). This *Blueprint* recognizes the critical role of risk analysis and suggests that available AI developers should identify and mitigate all risks, prioritizing "high impact risks" with proportionate mitigation, and suggests there be a clear role for independent evaluators to validate risks.[127] These evaluators might include researchers, journalists, ethics review boards, an inspector general, or certified third-party auditors. [128]

This section focuses on several key questions as shown in Table 1 for ADS safety, data privacy, and human-machine interaction issues. These are not the only questions pertaining to these topics but are included here to briefly frame the topic and think of strategies to move forward.

Table 1. Policy Considerations for ADS Safety

Blueprint Topics		Policy Approaches/Alternatives Raised
General ADS Safety	***Policy Options to Further Reform of FMVSS***	• USDOT to update the remaining 100 and 200- series FMVSS. • USDOT to reform exemption caps and/or exemption process. • USDOT to develop new ADS- specific FMVSS(s) oriented around risk analysis. • Invest in CARLA and other simulation tools to ensure they effectively measure safety across ADSs. • Create a standardized scenario catalog for simulation that differentiates between norms in different ODDs. This might include systems that can compare ADS performance to human drivers. • USDOT to establish an ADS Advisory Council to improve oversight (as proposed in SELF-DRIVE legislation) to provide additional oversight capacity.
	Policy Options to Advance a Safety Case Approach	• Establish guidance or requirements for ADS manufacturers to develop and maintain a Safety Case either in lieu of or in addition to FMVSS compliance. • Build accountability measures for ensuring risks are effectively assessed by employing independent oversight bodies to evaluate Safety Case submissions and make recommendations to decision makers.

[127] ID. AT 18.
[128] ID. AT 20.

Blueprint Topics		Policy Approaches/Alternatives Raised
		• Establish triggering events (e.g., fleet scaling, etc.) that would constitute the need for more specific risk-revaluation.

Data Collection and Privacy	***General Data Collection Policy Options***	• A federal ADS data exchange to evaluate ADS performance, overseen by an independent evaluation body, to collect:
		  ○ *leading metrics* - include hard breaking incidents, near-misses, unplanned stops, or instances of vehicles blocking roadways, traffic zones, or driveways.[129]
		  ○ *lagging metrics*, e.g., observable failure events, like accidents, injuries, and fatalities (these are collected currently but not housed in a comprehensive platform)
		  ○ Incident investigation boards that can evaluate these data holistically and make recommendations.
		• State data collection will focus on:
		  ○ passenger service - safety, sustainability, equity.
		  ○ safety driver compliance and behavior.
		  ○ remote operators' compliance and behavior.

Data Collection and Privacy	*Policy Options to Address Privacy Issues*	*Consumer Privacy* • Consider requirements for privacy risk assessment (as FMVSS or as an element of a Safety Case) that uphold the *Carpenter* ruling that the data is voluntarily provided, not considered indispensable to participation in modern society, nor does it constitute tracking over an extended time period. [130] • Establish tests for data security to ensure consumer data is held securely and held discrete from personal demographic information. [131] • Set guidance or policy for minimizing data collection and retention. *Proprietary Interests* • Establish clear criteria for what data constitutes a trade secret. • Establish protection protocols for preserving the integrity of any trade secret information that are determined to be necessary to collect.
Human Alternatives, Consideration, and Fallback	**Policy Options for Safety Drivers**	• Safety driver training and certification. • Driver legal liability awareness. • Fatigue risk management policies or shift lengths. • Alerts to address *automation complacency.* • Testing for *automation complacency.*

[129] BLUMENTHAL ET AL., *SUPRA* NOTE 6.
[130] *CARPENTER V. UNITED STATES*, SUPRA NOTE 104 AT 2220.
[131] THE PRIVACY IMPLICATIONS OF AUTONOMOUS VEHICLES, DATA PROTECTION REPORT (JUL. 17, 2017), HTTPS://WWW.DATAPROTECTIONREPORT.COM/2017/07/THE-PRIVACY-IMPLICATIONS-OF-AUTONOMOUS-VEHICLES/.

Blueprint Topics		Policy Approaches/Alternatives Raised
	Policy Options for Remote Operation	In addition to the above: • Remote operation classifications. • Remote operator authority or control levels for activities. • Ensuring functionality of remote operation stations. • Format of information sent to the operators. • Latency requirements or requirements for proximate location of remote operators. • Risk evaluation rubric to assess remote operators/classification during fleet scaling.

III. A. Blueprint for General ADS Safety

To summarize the discourse around two broad strategic ADS Safety approaches we will investigate: 1) Further Reforms to the FMVSS, and 2) ADS Safety Case Approach, and there are opportunities for blending these two policy alternatives. Rather than suggesting a specific recommendation, the aim is to highlight the merits and challenges of these two approaches.

Any framework will have to tackle the considerable challenge of accounting for a lack of transparency in AI and ADS manufacturing and operation. There are two broad types of AI, *white-box analysis*, where there is transparency and AI code is explainable, and *black-box analysis*, where the actions of an AI are unexplainable because the decision processes are unavailable, [132] or uninterpretable. Questions of explainability and transparency speaks to some of the questions raised with regards to robotics liability (see Section II, Page 19). Fault may only be causally determined after incidents where investigations can theoretically demand full access to vehicle code, but even then, it still may not be clear where the error occurred. Code that is explainable to individuals outside of an organization may be a black-box for those outside the ADS development group, which will put informational control in the hands of those who have white-box access. These issues of interpretability make it difficult for regulators to act proactively, given the privacy issues and costs of pervasive oversight.

Whether a standards-based system or a more flexible safety case framework will do better to address these explainability issues is still an open question. However, regulators still must make a strategic question to advance ADS safety. NHTSA recognized the merits of a more systemic approach in a document in the 2020 soliciting feedback on a *Framework for Automated Driving System Safety*. This *Framework* document makes

[132] MAJA BRKAN & GREGORY BONNET, *LEGAL AND TECHNICAL FEASIBILITY OF THE GDPR'S QUEST FOR EXPLANATION OF ALGORITHMIC DECISIONS: OF BLACK BOXES, WHITE BOXES AND FATA MORGANAS*, 11 EUROPEAN JOURNAL OF RISK REGULATION PAGE 34 (18.

mention of UL 4600, stating, “Traditional FMVSS may not be suitable for addressing certain critical safety issues relating to aspects of the core safety functions of perception, planning, and control.”[133]

After an initial public comment period, this *Framework* document never advanced to the next stage of policy development. There were hundreds of letters of public comments submitted to the USDOT in response to this introduced framework. These comments included many comments from bicycle advocates who suggested a broader focus on external safety not just occupant safety. Comments from the National society of Professional Engineers underscored the need for independent evaluators, calling for a “third-party verification process.” [134]

Further Reforming the FMVSS

Despite a lack of movement on the Framework USDOT has advanced several ADS-related reforms of FMVSS, with several additional reforms proposed that may advance. This may reflect a choice to work within the current FMVSS architecture, rather than developing a new approach. A recent FMVSS reform includes a 2022 update to occupant protection standards. NHTSA updated this standard to provide all passengers the same occupant crashworthiness protections when riding in an ADS-equipped vehicle, and the new standards recognizes that “innovative designs” may no longer include a “driver seat”, but that all passengers should be sufficiently protected in the event of a crash.[135] This change recognizes the existing regulatory bias which treats the driver as the most common, and therefore highest priority seat for crashworthiness. It is not clear whether this change will advance a broader conversation about how crashworthiness standards might better protect all vehicle occupants.[136]

The longest standing USDOT ADS strategy has been one built on exempting innovations that cannot squarely fit the standards. ADS developers can seek a limited number of exemptions to the FMVSS. For example, in 2022 General Motors (GM) sought a series of exemptions for their purpose-built Cruise origin vehicle, including FMVSS #104, windshield wiping, FMVSS #111, rear visibility, and FMVSS #208, occupant crash protection, among others.[137] While there are clear reasons why a vehicle without a driver does not need rear visibility in the same way that driven vehicles do, concerns have been raised about whether this exemption approach results in a misalignment of overall Federal agency objectives. Legal scholars Sheriff and Gossett also point out, “NHTSA's obsession with 'removing regulatory hurdles' and 'not stifling innovation' is inconsistent

[133] FRAMEWORK FOR AUTOMATED DRIVING SYSTEM SAFETY, *SUPRA* NOTE 129.
[134] SHERIFF KATEHERINE & DAVID GOSSETT, *AUTOMATED DRIVING REGULATION: THREE BIG ISSUES AS THE FEDERAL GOVERNMENT EDGES FORWARD*, ARTIFICIAL INTELLIGENCE LAW ADVISOR (MAR. 18, 2021), HTTPS://WWW.DWT.COM/BLOGS/ARTIFICIAL-INTELLIGENCE-LAW-ADVISOR/2021/03/DOT-NHTSA-AUTOMATED-DRIVING-SYSTEM-REGULATIONS.
[135] OCCUPANT PROTECTION FOR VEHICLES WITH AUTOMATED DRIVING SYSTEMS, (2022).
[136] FRAMEWORK FOR AUTOMATED DRIVING SYSTEM SAFETY, *SUPRA* NOTE 129.
[137] GENERAL MOTORS-RECEIPT OF PETITION FOR TEMPORARY EXEMPTION FROM VARIOUS REQUIREMENTS OF THE FEDERAL MOTOR VEHICLE SAFETY STANDARDS FOR AN AUTOMATED DRIVING SYSTEM-EQUIPPED VEHICLE, FEDERAL REGISTER (2022), HTTPS://WWW.FEDERALREGISTER.GOV/DOCUMENTS/2022/07/21/2022-15557/GENERAL-MOTORS-RECEIPT-OF-PETITION-FOR-TEMPORARY-EXEMPTION-FROM-VARIOUS-REQUIREMENTS-OF-THE-FEDERAL (LAST VISITED JUN 8, 2023).

with the agency's mission 'to save lives, prevent injuries, and reduce economic costs due to road traffic crashes."[138]

On the other hand, there have been others who are calling for increases in the number of allowable ADS FMVSS exemptions. Federal legislative discussions regarding reforming the FMVSS exemption process for ADS manufacturers has been largely focused on increasing exemption caps. Recent 2023 legislation proposed raising the cap from 2,500 exempted vehicles per manufacturer to 25,000 in the first 12-month period after enactment, 50,000 the next, and 100,000 for the 3rd and 4th. To contextualize these exemptions cap increases, this would be a significant scaling, as high as a 40% increase over the 4-years.[139]

Several other reforms to FMVSS have been proposed, including those listed in a 2020 USDOT report summarizing 100- and 200-series FMVSS that may need reform. These include reforms for how to ensure safety of bi-directional capabilities. [140]

Role of Simulation Software

Testing ADS in real-world scenarios is necessary but it is extremely costly, so the role of simulations will be essential to train and test any ADS. The USDOT has made an open-source AD simulator available for testing any ADS. The first version was released in 2017, referred to as the Car Learning to Act (CARLA) allows automated driving software to be tested in different environmental settings, using different map generators. CARLA is highly customizable and can mimic ADS sensors to test:

> *CARLA is grounded on Unreal Engine to run the simulation and uses the OpenDRIVE standard to define roads and urban settings. Control over the simulation is granted through an application programming interface (API) handled in Python® and C++.*[141]

The EU funded the Safety Assurance Framework for Connected, Automated Mobility Systems (SUNRISE) project has sought to create a standardized scenario catalog for testing and aims to break down siloing and enable collaboration in a harmonized way between regulators, developers, industry, and the public.[142] Japan's JAMA is pursuing a similar goal.[143] In the U.S., there has been some interest in identifying common scenarios, for example the SAE International On road Automated Driving (ORAD) committee has set out to investigate Best Practices for Developing and Validating Simulations for Automated Driving Systems, but (as of December 2023) the committee had been unable to come to any consensus on safety metrics, assessment methods, or

---

[138] KATEHERINE AND DAVID GOSSETT, *SUPRA* NOTE 130.
[139] LAURA FRAADE-BLANAR & NIDHI KALRA, *AUTONOMOUS VEHICLES AND FEDERAL SAFETY STANDARDS: AN EXEMPTION TO THE RULE?*, (2017), HTTPS://WWW.RAND.ORG/CONTENT/DAM/RAND/PUBS/PERSPECTIVES/PE200/PE258/RAND_PE258.PDF.
[140] NATIONAL HIGHWAY TRAFFIC SAFETY ADMINISTRATION, *FMVSS CONSIDERATIONS FOR VEHICLES WITH AUTOMATED DRIVING SYSTEMS: VOLUME 1*, 1 (2020), HTTPS://WWW.NHTSA.GOV/SITES/NHTSA.GOV/FILES/DOCUMENTS/ADS-DV_FMVSS_VOL1-042320-V8-TAG.PDF.
[141] INTRODUCTION - CARLA SIMULATOR, HTTPS://CARLA.READTHEDOCS.IO/EN/LATEST/START_INTRODUCTION/.
[142] ABOUT | SUNRISE PROJECT, HTTPS://CCAM-SUNRISE-PROJECT.EU/ABOUT/ (LAST VISITED JUL 14, 2023).
[143] JAPAN AUTOMOBILE MANUFACTURERS ASSOCIATION INC, JAMA - JAPAN AUTOMOBILE MANUFACTURERS ASSOCIATION, INC, JAMA - JAPAN AUTOMOBILE MANUFACTURERS ASSOCIATION, INC, HTTPS://WWW.JAMA.OR.JP (LAST VISITED JUL 25, 2023).

validation approaches.[144] This may reflect apprehension from private industry to develop common simulation tools or standards for simulation. Industry parties are vocal about the challenges of standardizing simulations—given that simulation capabilities are a core part of all AV operations, there are varied and private considerations. Developers have raised cautions about whether a standard simulation could apply across development stacks.

In Germany, the PEGASUS project, promoted by the German Federal Ministry for Economic Affairs and Energy, has sought to develop a liability minimizing database by creating a baseline set of scenarios for an AV to test against. This would create a standardized approach to verifying and validating ADS safety and help determine shortcomings in automated systems.[145] In the UK a similar project is being pursued by the University of Warwick and Deepen AI called the Safety Pool Initiative, seeking to create a global safety assessment framework for ADSs built off a curated Safety Pool Scenario Database.[146]

FMVSS Reform Purview Policy Options

This policy will likely be led by the USDOT, with collaboration from other agencies, and states. Potential FMVSS reforms include:

- USDOT to update the known remaining 100 and 200- series FMVSS.
- USDOT to reform exemption caps and/or exemption process.
- USDOT to develop new ADS- specific FMVSS(s) oriented around risk analysis.
- Invest in CARLA and other simulation tools to ensure they effectively measure safety across ADSs.
- Create a standardized scenario catalog for simulation that differentiates between norms in different ODDs. This might include systems that can compare ADS performance to human drivers.
- USDOT to establish an ADS Advisory Council to improve oversight (as proposed in SELF-DRIVE legislation) to provide additional oversight capacity.

Safety Case Approach

A *safety case*, (as discussed in more detail in the Safety Background Section on Page 13) is an iterative framework that provides developers a method to input environmental factors, and update risks based on what is observed. While it can include standards, or benchmarks for cataloging hazards and risks, it is not meant to be a simple checklist of standards. Safety researcher, Koopman, is supportive of an iterative safety case approach, due to concerns that static vehicle-specific standards are insufficient for determining system-wide risks.

[144] HTTPS://STANDARDSWORKS.SAE.ORG/STANDARDS-COMMITTEES/ROAD-AUTOMATED-DRIVING-ORAD-COMMITTEE
[145] *PEGASUS METHOD: AN OVERVIEW*, HTTPS://WWW.PEGASUSPROJEKT.DE/FILES/TMPL/PEGASUS-ABSCHLUSSVERANSTALTUNG/PEGASUS-GESAMTMETHODE.PDF.
[146] SAFETY POOL - KEY MEMBERS, HTTPS://WWW.SAFETYPOOL.AI/ABOUT (LAST VISITED JUL 26, 2023).

Rather than adopting a fiction that mere conformance to a standard at deployment results in flawless risk mitigation, instead it is important to continually evaluate and improve the residual risk present in the system.[147]

A safety case is both an internal explanatory tool to coordinate across different engineering teams within an organization, and it is a tool for strategic external communication about what safety procedures have been considered. While the Safety Case has many advocates, critics also point to several inadequacies in the existing safety case approach:[148]

- Self-reported evidence and the proprietary nature of the claims made in a safety case make independent evaluation a challenge.
- The leading safety case standard, UL4600 is intentionally flexible, so lacks pass/fail criteria.[149]
- Standards organizations often lack a robust community engagement process and may reflect representational biases. The panels that approve many of the consensus standards typically reflect a consensus only of the members of the organizations participating in the panel (which costs considerable resources and time to participate) thus potentially overshadowing the voices of government, academic, or other stakeholders not engaged in these organizations.

Whether a state or federal version of a safety case approach can address these criticisms is not evaluated here, but for this approach to advance, more iteration will likely be necessary to improve validity and accountability of the approach, while preserving flexibility. An effective safety case requires effective independent assessors.[150] This requirement aligns the safety case approach with a priority stated in the White House *Blueprint for an AI Bill of Rights*, which states, "Automated systems should be designed to allow for independent evaluation (e.g., via application programming interfaces)." The White House shares that evaluators should include researchers, journalists, ethics review boards, inspector general, or third-party auditors. This multi-disciplinary list of evaluators highlights the importance of making performance data available to establish and ensure ongoing accountability. [151]

Safety Case: Purview and Policy Options

States or federal regulators could pursue a Safety Case approach. However, USDOT rulemakings to establish a safety case framework would likely take several years to advance. In the interim, states may choose to adopt interpretations of safety case frameworks that align best with their state's policy objectives, and these reforms would also take time to refine. These reforms might include:

---

[147] PHILIP KOOPMAN ET AL., *A SAFETY STANDARD APPROACH FOR FULLY AUTONOMOUS VEHICLES*326, 5 (2019), HTTPS://USERS.ECE.CMU.EDU/~KOOPMAN/PUBS/KOOPMAN19_WAISE_UL4600.PDF.
[148] CAMILA CORREA-JULLIAN, JOACHIM GRIMSTAD, SPENCER AUGUST DUGAN, MARILIA RAMOS, CHRISTOPH A. THIEME, ANDREY MOROZOV, INGRID B. UTNE, AND ALI MOSLEH. PROCEEDINGS OF THE 4TH INTERNATIONAL WORKSHOP ON AUTONOMOUS SYSTEMS SAFETY (IWASS 2023). DOI:10.34948/G4MW2N.
[149] ID.
[150] KOOPMAN ET AL., *SUPRA* NOTE 144.
[151] ID. AT 20.

- Establishing guidance or requirements for ADS manufacturers to develop and maintain a holistic risk management plan, or safety case either in lieu of or in addition to FMVSS compliance.
- Building accountability measures for ensuring risks are effectively assessed by employing independent oversight bodies. These entities can evaluate safety case submissions, which would enable them to serve as sources of verifiable evidence in future claims and make recommendations to decision makers on the safety case conclusions.
- Establishing triggering events (e.g., fleet scaling) that would constitute the need for more specific risk-revaluation.

III. B. Blueprint for ADS Data Collection and privacy

This section will highlight key lessons in public documents and the literature on data collection and data privacy for ADS. ADS data is just a tip of a large iceberg of data collection occurring across the mobility and consumer electronics sector. It is possible that some of the issues raised in this section could be resolved through broader AI policy, however, to the extent possible, this section will focus on issues that are unique to ADS privacy, or differentiate ADS equipped vehicles.

The dependence of ADS-equipped vehicles on external sensors to perceive and interpret the world and make informed decisions requires gathering massive amounts of training data to ensure that the models in use are working effectively. Ongoing oversight and additional guidance will ensure that each ADS fleet can scale safely, grow competence, and demonstrate continuous improvement. But regulators will have to establish what data is necessary, what data format, how often the data is updated, and how to preserve privacy and proprietary interests.

General Data Collection Considerations

The Automated Vehicle Safety Consortium offers a possible framework for the continuous monitoring and improvement of performance as known and unknown changes arise in the operating environment.[152] This Consortium suggests, "[c]ertain trends or anomalies may indicate changes have occurred within the ADS-DV's operating environment or inaccuracies in assumptions made prior to deployment."[153] By establishing performance metrics that can flag where irregularities emerge as compared to previous expectations or assumptions, developers can adjust these assumptions if needed and develop, test, and implement appropriate responses for their vehicles.[154]

[152] AUTOMATED VEHICLE SAFETY CONSORTIUM. 2023. *AVSC BEST PRACTICE FOR CONTINUOUS MONITORING AND IMPROVEMENT AFTER DEPLOYMENT*. SAE INDUSTRY TECHNOLOGIES CONSORTIA.
[153] ID. AT 7.
[154] ID. AT 7-10.

In some ways, this wait-and-see-where-anomalies-emerge stance aligns with the status quo of the U.S. federal approach to safety oversight. Only when anomalies present as accidents or observable safety failures, are driven vehicles required to respond to more detailed data inquiries to investigate issues, potentially face voluntary recalls, and forced to identify suggestions for remedies. This may point to a reactive type of data collection, focused on lagging metrics, as a more likely outcome for the U.S. But it is unclear whether this approach will prove sufficient to assure the public that ADSs are safe. Different stakeholders may have different positions that reflect a blend of policy goals and norms (e.g. federal, state, or local) that favor a more permissive, or reactive approach and some favoring a more controlling or proactive approach.[155]

Also, it is not straightforward to determine whether and how to take a proactive or reactive approach. Leading metrics may be possible forewarnings of more serious potential failures on the horizon, in practice it will be difficult to assess whether these early warning signs are warnings after all, and an excessive focus on specific leading measures may harm safety outcomes. For example, early data monitoring reporting focused on disengagements (of the automated system), where safety drivers or operators intervened during testing. This leading metric has been widely criticized for not fully capturing or reflecting the nature of progress in the testing process. Reporting disengagements may penalize companies that disengage more frequently out of an abundance of caution, while rewarding those that delay disengagement where they may have been necessary. This example can be a lesson in how to ensure that leading metrics are revised when there is evidence that the metric is resulting in unintended consequences.

These issues may point to the need for more holistic assessments, rather than looking at specific data, or waiting for a specific component to fail. A more holistic assessment can establish better metrics based on combined objectives (e.g., compliance with FMVSS, functional success/failure) and determine how to minimize risk within these parameters.

For either approach to work, data collection will be critical, and several notable examples may serve as a model for ADS data collection:

- **USDOT data exchanges**: The USDOT Secure Data Commons (SDC) is an example of how to collect and manage private data securely. The platform enables vehicle data to be uploaded periodically, or in real-time, and users have tiered access, some could use analytical tools in R but not download raw data, while others may be granted more access.[156]

- **The Mobility Data Specification (MDS):** Over 300+ global cities are using an open-source data standard, originating out of the City of Los Angeles, overseen by a non-profit standards-setting organization, the Open Mobility Association. MDS facilitates real time data exchange between regulators and private actors and incorporate real-time and historic data exchange. Historically the

[155] KELLY FLEMING, *TECHNOLOGY IS OUTPACING STATE AUTOMATED VEHICLE POLICY* (2020), HTTPS://ESCHOLARSHIP.ORG/UC/ITEM/0K85R9JV.
[156] SECURE DATA COMMONS - SDC LIFECYCLE | US DEPARTMENT OF TRANSPORTATION, HTTPS://WWW.TRANSPORTATION.GOV/DATA/SECURE/SDC-LIFECYCLE (LAST VISITED APR 23, 2024).

MDS data specification has focused on scooter vehicles.[157] Trip level data must be held securely by the city or a contracted third party.[158] MDS also enables two-way data sharing where regulators and cities can add data to the API that can be usable by private operators, and operators also can add data via an application programming interface (API).[159] Many cities using MDS, including Los Angeles, mandate the use of MDS before approving e-scooter permits. While initially developed for micromobility, such as e-scooters and bikeshare, the Open Mobility Foundation expanded the MDS 2.0 to enable data sharing for ride-hailing, taxi, and other passenger services, so questions remain about how to augment the MDS or whether creation of a new specification for ADS is necessary.[160]

- **Federal Aviation Administration Oversight.** FAA's risk-based oversight process, composed of information sharing and reporting programs is attributed to many lives saved.[161] These programs include the Aviation Safety Information and Sharing (ASIAS) program, voluntary reporting programs, and Aviation Safety Infoshare. The ASIAS data repository includes public and proprietary data sources, and brings together government and industry sources.[162] The Commercial Aviation Safety Team detects risks and implements mitigation strategies based largely on the voluntary reporting programs and working closely with the ASIAS program.[163] [164] FAA inspectors check to ensure carrier Safety Management System processes and consider these risks along with those identified by ASIAS system. If they fail to do so, the FAA can require increased surveillance through the Safety Assurance System. Commercial Aviation Safety Team, ASIAS, the FAA along with industry, participate in the semi-annual Aviation Safety Infoshare where they share safety concerns and best practices under confidential protections in a non-punitive reporting environment. The FAA's success shows the potential for collaborative information sharing programs between industry and government along with tools that AV regulators can use to facilitate open data sharing.

ADS Data: Purview and Policy

[157] ABOUT MDS | OPEN MOBILITY FOUNDATION, (2020), HTTPS://WWW.OPENMOBILITYFOUNDATION.ORG/ABOUT-MDS/ (LAST VISITED JUN 15, 2023).
[158] *RE-IDENTIFICATION OF "ANONYMIZED" DATA*, GEORGETOWN LAW TECHNOLOGY REVIEW (2017), HTTPS://GEORGETOWNLAWTECHREVIEW.ORG/RE-IDENTIFICATION-OF-ANONYMIZED-DATA/GLTR-04-2017/ (LAST VISITED JUN 15, 2023).
[159] OPEN MOBILITY FOUNDATION, HTTPS://WWW.OPENMOBILITYFOUNDATION.ORG/ (LAST VISITED JUN 15, 2023).
[160] HTTPS://GITHUB.COM/OPENMOBILITYFOUNDATION/GOVERNANCE/RAW/MAIN/DOCUMENTS/OMF-MDS-ARCHITECTURAL-LANDSCAPE.PDF
[161] OUT FRONT ON AIRLINE SAFETY: TWO DECADES OF CONTINUOUS EVOLUTION | FEDERAL AVIATION ADMINISTRATION, HTTPS://WWW.FAA.GOV/NEWSROOM/OUT-FRONT-AIRLINE-SAFETY-TWO-DECADES-CONTINUOUS-EVOLUTION (LAST VISITED JUL 26, 2023).
[162] AVIATION SAFETY INFORMATION ANALYSIS AND SHARING PROGRAM | FEDERAL AVIATION ADMINISTRATION, HTTPS://WWW.FAA.GOV/NEWSROOM/AVIATION-SAFETY-INFORMATION-ANALYSIS-AND-SHARING-PROGRAM-1 (LAST VISITED JUL 26, 2023).
[163] ID.
[164] OUT FRONT ON AIRLINE SAFETY: TWO DECADES OF CONTINUOUS EVOLUTION | FEDERAL AVIATION ADMINISTRATION, SUPRA NOTE 156.

A purely federally led data collection effort may result in gaps for many states and cities.[165] State and federal governance for monitoring ADS safety is shared, but who should be collecting what information will be dependent on legislative direction and alignment with different policy objectives. Duplicative data collection is occurring, and there are calls for a shared API or common data specification to improve compatibility and enable a centralized data collection system where states can easily add criteria. This might address different data needs across different governmental authorities. Policy alternatives to support monitoring and evaluation of ADS systems might include:

- A federal ADS data exchange to evaluate ADS performance, overseen by an independent evaluation body, to collect:
    - *leading metrics* - include hard breaking incidents, near-misses, unplanned stops, or instances of vehicles blocking roadways, traffic zones, or driveways.[166]
    - *lagging metrics*, e.g., observable failure events, like accidents, injuries, and fatalities (these are collected currently but not housed in a comprehensive platform.
    - Incident investigation boards that can evaluate these data holistically and make recommendations.
- State data collection will focus on:
    - passenger service - safety, sustainability, equity.
    - safety driver compliance and behavior.
    - remote operators' compliance and behavior.

ADS Data Privacy Considerations

There is widespread consensus that data collected by mobility companies has value and utility within its intended primary use, and outside of that use. Xie et al. point out that,

> *…training AVs usually requires a large amount of training data collected from different driving environments (e.g., cities) as well as different types of personal information (e.g., working hours and routes). Such collected large data, treated as the "new oil" for ML in the data-centric AI era, usually contains a large amount of privacy sensitive information which is hard to remove or even audit."* [167]

[165] MOLLIE D'AGOSTINO, PAIGE PELLATON & AUSTIN BROWN, MOBILITY DATA SHARING: CHALLENGES AND POLICY RECOMMENDATIONS, (2019), HTTPS://ESCHOLARSHIP.ORG/UC/ITEM/4GW8G9MS.
[166] BLUMENTHAL ET AL., SUPRA NOTE 6.
[167] XIE ET AL., *SUPRA* NOTE 8.

Protecting this sensitive information will be a balancing act. Both oversharing and under-sharing data can raise known and unknown risks for different parties.[168] This section will simplify this broad topic by narrowing the focus to two domains: consumer privacy and commercial proprietary privacy.

Consumer privacy risks

Consumers of mobility services typically grant access to many potential sensitive data points, which are held by ADS operators according to user agreements. These might include movement and route data, video, or images of individuals inside the vehicles, as well as personal information regarding riders age, gender, or other characteristics. [169] This type of data gathering is not unique to an ADS setting, and similar data are collected by ridehailing fleet operators, or to a lesser extent some OEMs, especially those with ADAS features. However, the potential for capture of in-vehicle videos and images will likely add an additional layer of sensitivity.

The status quo for automotive manufacturers is to control and tightly manage large stores of data for millions of vehicles in their fleets. For example, manufacturers have vehicle event data recorders (EDRs), commonly referred to as black boxes, that can help assess vehicle faults and remedies by storing data on vehicle modes and actions in the moments before airbags deployments. Based on Driver Privacy Act of 2015,[170] this data can only be downloaded with owner consent to repair the vehicle, or if mandated by an official investigation (by regulator or court official).

Certain data risks to consumers are known:

- Human trace data: As was discussed in the legal privacy Section, route data over an extended period is easily identifiable given people's typical movement patterns (e.g., home-to-work, home-to-shopping destinations, etc.). Reidentification, public exposure of sensitive information, targeted cyber-criminal activity, or the most severe outcome is stalking, or violence perpetrated on individuals due to data exposure.

- Personal demographic information: such as age, ethnicity, etc., can be used to "…harass AV users through marketing and advertising, to steal users' identity, profile users and predict their actions, concentrating information and power." 171

- Remote surveillance and video or photographic information: Some ADS operation will have live video feeds to ensure compliance with rules of operation, and to monitor the vehicles, but these feeds would be highly sensitive, and will represent a new opportunity for access to activities occurring in transport, which has implication for both governmental access (law enforcement, legal proceedings) and criminal

[168] D'AGOSTINO, PELLATON & BROWN, *SUPRA NOTE 162.*
[169] XIE ET AL., *SUPRA* NOTE 8.
[170] SEN. JOHN HOEVEN, *DRIVER PRIVACY ACT OF 2015*, TITLE 49 U.S. CODE (2015), HTTPS://WWW.CONGRESS.GOV/BILL/114TH-CONGRESS/SENATE-BILL/766/TEXT.
[171] ARAZ TAEIHAGH & HAZEL SI MIN LIM, GOVERNING AUTONOMOUS VEHICLES: EMERGING RESPONSES FOR SAFETY, LIABILITY, PRIVACY, CYBERSECURITY, AND INDUSTRY RISKS, 39 TRANSPORT REVIEWS 103 (2019).

wrong-doers (ransom, hijacking, etc.). Whether ADS vehicles are considered a public space will dictate what privacy protections riders will retain. [172]

Consumer Data Privacy: Purview and Policy Options

While purview over privacy laws will be shared by federal authorities, states and other jurisdictions, coordination will be key to ensure that privacy is preserved across both consumers and commercial interests.

- Consider requirements for privacy risk assessment (as FMVSS or as an element of a Safety Case) that uphold the *Carpenter* ruling that the data is voluntarily provided, that using this transportation services is not considered indispensable to participation in modern society, nor does it constitute tracking over an extended time period.[173]
- Establish tests for data security to ensure consumer data is held securely and held discrete from personal demographic information. [174]
- Set guidance or policy for minimizing data collection and retention.

Commercial Privacy Risks

Automotive manufacturers control and securely manage large stores of data. For example, vehicles have event data recorders (EDRs) that assess vehicle faults and remedies by storing data for the moments before airbags deploy. There are many other types of data held by OEMs and ADS developers. Whether these data are considered a trade secret will rest on policy and court determinations. In a 2022 court case involving Waymo and the State of California, where data collected were shared with a journalist via a public records request, the judgement stated that "operational processes and design capabilities"[175] do constitute a trade secret that can be held for some period by the state, but this protection was not preserved indefinitely.[176] This can be interpreted to mean, that if trade secret data is held it may not be sharable for some period of time. States already hold securely some data on ADS operations. However, questions remain on what circumstances justify an ADS developer to provide information classified as a trade secret to regulatory authorities, and what protections these organizations will retain.

Commercial Data Privacy: Purview and Policy Options

Both state and federal authorities and courts continue to weigh in on commercials privacy. The following are considerations for this policy area:

- Establish clear criteria for what data constitutes a trade secret.

---

[172] ID.
[173] CARPENTER V. UNITED STATES, SUPRA NOTE 105 AT 2220.
[174] THE PRIVACY IMPLICATIONS OF AUTONOMOUS VEHICLES, *SUPRA* NOTE 128.
[175] OPENING BRIEF FOR PLAINTIFF, WAYMO LLC, V. CAL. DMV, NO. 34-2022-80003805-CU-WM-GDS (CAL. SUPER. CT. FILED FEB. 2, 2022), 7.
[176] OPENING BRIEF FOR PLAINTIFF, WAYMO LLC, V. CAL. DMV, NO. 34-2022-80003805-CU-WM-GDS (CAL. SUPER. CT. FILED FEB. 2, 2022), 7.

- Establish protection protocols for preserving the integrity of any trade secret information that are determined to be necessary to collect.

III. C. Blueprint for ADS Human Alternatives, Consideration, and Fallback

Humans are likely to remain involved in some aspects of ADS operational safety and service. Two key human roles that may require policy attention are the *remote operator* and the *safety driver*. If ADS systems rely on one or both roles, these roles must be integrated into the risk management efforts. It may be difficult for regulators to accurately validate human error and control issues. Whether human errors initiate or contribute to accidents will be situationally specific. For example, humans in supervisory roles can fail to react to hazardous events resulting in *human-out-of-loop* issues that can lead to accidents or missteps. However, effective supervision and well-timed intervention can yield safer operations. Automation that is too reliant on operators may pose challenges regarding human-system interaction safety, given what is known about challenges in this area.[177]

These issues present complexities for both worker protocols and legal obligations. These roles will vary considerably across ADS developers, and thus, there may be a need for more concrete standards about how to evolve the safety driver and remote operator roles equitably and safely, to ensure that human ADS operators can be most successful.

Considerations for Safety Drivers

Safety drivers are often a critical element in testing the safety of most ADS-equipped passenger vehicles that can accommodate such roles. For some ADS fleets, safety drivers may have a longer-term role. For example, ADS vehicles for use in publicly supported settings, e.g. May Mobility microtransit vehicles fleets, have kept some on-board safety drivers or safety supervisors to assist certain types of passengers and routes while removing safety drivers from other routes.[178] Another example, Pony.ai safety drivers in 2021 were reportedly not responsible for steering or otherwise maneuvering the vehicle, but they had "a red button that can stop the vehicle just in case anything happens."[179]

Staff or technicians may also be deployed on an ad-hoc basis. If a vehicle has stopped to achieve a minimal risk condition, or if there is an accident, humans may assist with retrieval and/or removal of the ADS vehicle. In some cases, these drivers will be necessary to move or drive the vehicle manually. ADS vehicles without any on-

[177] MARILLIA A. RAMOS ET AL., *ACCOUNTING FOR HUMAN FAILURE IN AUTONOMOUS SHIP OPERATIONS*, *IN* SAFETY AND RELIABILITY – SAFE SOCIETIES IN A CHANGING WORLD 355 (2018), HTTPS://WWW.TAYLORFRANCIS.COM/CHAPTERS/OA-EDIT/10.1201/9781351174664-45/ACCOUNTING-HUMAN-FAILURE-AUTONOMOUS-SHIP-OPERATIONS-RAMOS-UTNE-VINNEM-MOSLEH.

[178] ED GARSTEN, *MAY MOBILITY TAKES SAFETY DRIVERS OUT OF ROBO TRANSITS IN SUN CITY*, FORBES, DEC. 18, 2023, HTTPS://WWW.FORBES.COM/SITES/EDGARSTEN/2023/12/18/MAY-MOBILITY-TAKES-SAFETY-DRIVERS-OUT-OF-ROBO-TRANSITS-IN-SUN-CITY/.

[179] HYUNJOO JIN, *INSIGHT: A SECRET WEAPON FOR SELF-DRIVING CAR STARTUPS: HUMANS | REUTERS*, REUTERS, AUG. 23, 2021, HTTPS://WWW.REUTERS.COM/BUSINESS/AUTOS-TRANSPORTATION/SECRET-WEAPON-SELF-DRIVING-CAR-STARTUPS-HUMANS-2021-08-23/.

board operational capacity, such as traditional steering wheels and brakes may have other means of stopping the vehicle (e.g., on-board panels or controls). These vehicles may still require technicians to be available to reset, manually assist, arrange towing, etc.

Regulatory requirements for the presence of safety drivers vary, but some U.S. states require safety drivers during specific testing phases, and/or in the event of an incident. In California, ADS operators typically use safety drivers in early testing phases, but they can apply for a permit to deploy without first testing with a safety driver permit. Removing an in-vehicle human usually is a process whereby the ADS training can demonstrate it can operate safely or provide sufficient evidence of the absence of unreasonable risk.

Safety drivers will, in many cases, also be held legally responsible for that safe operation, especially if there are collisions when they are operating using traditional steering wheels and brakes. The Vasquez criminal case (see the Defining Safety: Legal Liability section for more details) is evidence that ADS safety drivers are legally responsible for driving tasks. They are obligated to be alert, attentive, and maintain situational awareness of the vehicle, in any roadway conditions or circumstances that they encounter. It may require policy to ensure that safety drivers are well-aware and informed of their legal obligations and the criminal consequences that may accompany their negligent actions.

Safety Drivers: Purview and Policy Options

Several policy alternatives may be available with respect to safety drivers. Federal and/or state authorities might collaborate to establish guidance or set standards for the following safety driver metrics (although further evaluation is necessary to determine impacts of these interventions):

- Safety driving training guidance or standards for testing.
- Driver statements recognizing their legal liability while performing their duties.
- Fatigue risk management policies.
- Limits of shift lengths.
- Auditory or visual alerts or testing to mitigate incidents of *automation complacency*.

Considerations for Remote Operators and Remote Operation

Safety researcher Kalaiyarasan defines *remote operation* as,

> *…an umbrella term that encompasses the functions needed to support the operations of an AV or a fleet of AVs by a remote operator. Remote operation might include both driving and non-driving related tasks.*[180]

There are many examples of different remote operational classifications deployed by different ADS developers. Remote service assistance, where an operator can communicate directly with passengers will likely be pervasive. However, other types of remote operation will likely vary across the ADS industry. For example,

[180] ARUN KALAIYARASAN, *REMOTE OPERATION OF CAVS - THE NEED FOR DEFINITION*, TRL: THE FUTURE OF TRANSPORT (2021), HTTPS://TRL.CO.UK/NEWS/REMOTE-OPERATION-OF-CAVS---THE-NEED-FOR-DEFINITION.

Waymo reported in 2020 that they were not using any fully remote operators at that time.[181] However, while they do not use "fully remote operations" the company did intervene with remote assistance. For example, in March 2024 a Waymo ADS vehicle entered an obstructed intersection with construction against a red light, after receiving erroneous remote guidance to advance into the intersection. Remote assistance is also a strategy used by Cruise ADS vehicles. In a New York Times article in November 2023 while ADS developer, Cruise was operating around 400 vehicles in San Francisco, remote interventions occurred periodically, "workers intervened to assist the company's vehicles every 2.5 to five miles."[182] However, this statistic may have been taken out of adequate context. Former Cruise CEO, Kyle Vogt, stated that this reporting was referring to all interventions, including those that are ultimately resolved by the ADS as well as those that can be resolved within a second. Vogt stated that, "Cruise AVs are being remotely assisted (RA) 2-4% of the time on average, in complex urban environments."[183] These discrepancies between media accounts of remote operator intervention and reports from the primary sources highlight an information validity gap, and a lack of clear definitions for remote operations or assistance services, which may contribute to public distrust.

Other ADS operators are not shying away from their use of remote operation. Vay, a Berlin based company started in 2018, has leveraged teleoperation to get their cars on the road in Germany. Their approach to automated deployment is to teleoperate vehicles enroute to users who request a ride, then have consumers drive themselves where they need to have a teleoperator drive the car back. Their cars started driving in Hamburg in February 2022 without a safety driver in the car, solely controlled by a teleoperator.[184]

Remote Operator/Operations: Purview and Policy Options

Remote operators, remote operation, or remote assistance require more comprehensive definitions. These options can also only be supported by good connectivity (latency requirements), actionable information (what information is shared and when), and effective workplace design and training materials (e.g., ratio of operators to vehicles monitored simultaneously, effective training on how intervention alarms work, etc.).

Regarding regulatory purview in the U.S. remote operation and remote operators are divided. Remote operation will be managed by each ADS, and therefore federal vehicle regulatory requirements could be designed to ensure existence or set certain specifications for remote operating and remote assistance. However, there are some grey areas on purview because state DMVs also see ensuring safe vehicle operation as part of their mandate. It is clearer that states would manage any licensure or testing requirements for

---

[181] THE WAYMO TEAM, *SHARING OUR SAFETY FRAMEWORK FOR FULLY AUTONOMOUS OPERATIONS*, WAYPOINT (OCT. 30, 2020), HTTPS://WAYMO.COM/BLOG/2020/10/SHARING-OUR-SAFETY-FRAMEWORK/.

[182] TRIPP MICKLE, CADE METZ & YIWEN LU, *G.M.'S CRUISE MOVED FAST IN THE DRIVERLESS RACE. IT GOT UGLY.*, NEW YORK TIMES, NOV. 3, 2023, HTTPS://WWW.NYTIMES.COM/2023/11/03/TECHNOLOGY/CRUISE-GENERAL-MOTORS-SELF-DRIVING-CARS.HTML#:~:TEXT=THE%20WORKERS%20INTERVENED%20TO%20ASSIST,THAT%20IT%20WAS%20HAVING%20PROBLEMS.

[183] KVOGT, *CRUISE CEO HERE. SOME RELEVANT CONTEXT FOLLOWS.*, HACKER NEWS (2023), HTTPS://NEWS.YCOMBINATOR.COM/ITEM?ID=38145997#:~:TEXT=%3E%3E%20CRUISE%20AVS%20ARE%20BEING%20REMOTELY,REVIEW%20THINGS%20IN%20CERTAIN%20SITUATIONS.

[184] VAY, *A HISTORIC MOMENT: THE FIRST CAR DRIVES WITHOUT A PERSON IN THE VEHICLE ON A EUROPEAN PUBLIC ROAD - VAY*, (2023), HTTPS://VAY.IO/PRESS-RELEASE/A-HISTORIC-MOMENT-THE-FIRST-CAR-DRIVES-WITHOUT-A-PERSON-IN-THE-VEHICLE-ON-A-EUROPEAN-PUBLIC-ROAD.

remote operators. Thus far, California is one state requiring the existence of a remote operator, but there are not detailed specifications regarding this role.[185]

In addition to the human-system interaction elements discussed in relation to safety driver, several remote assistance policy alternatives are available. Those listed below will require further evaluation to assess how can safety improvements or other public objectives can be achieved.

- Remote operation classifications pertaining to the activities operators undertake, e.g., remote user assistance, remote vehicle assistance, and remote control, and identify what is reasonably expected of each of these roles to achieve ADS safety and worker protections.
- Remote assistance and control authority or level of control for different activities.
- Format of information sent to the operators, to give sufficient situational awareness and transparency but also to avoid information overload.[186]
- Latency requirements or requirements for proximate location of remote operators (e.g. requiring domestic, or regional locations) with respect to safety and workforce priorities.
- Ensuring the functionality of manual controls for remote operation stations, based on each of the remote classifications.
- Risk evaluation rubric to assess how the number of remote operators in each classification should change with each marginal vehicle in active deployment.

[185] CALIFORNIA AUTONOMOUS VEHICLE REGULATIONS, CALIFORNIA DMV, HTTPS://WWW.DMV.CA.GOV/PORTAL/VEHICLE-INDUSTRY-SERVICES/AUTONOMOUS-VEHICLES/CALIFORNIA-AUTONOMOUS-VEHICLE-REGULATIONS/
[186] ARUN KALAIYARASAN, *REMOTE OPERATION OF CAVS - THE NEED FOR DEFINITION*, TRL: THE FUTURE OF TRANSPORT (2021), HTTPS://TRL.CO.UK/NEWS/REMOTE-OPERATION-OF-CAVS---THE-NEED-FOR-DEFINITION.

VI. Conclusions and Recommendations for Further Research

Key takeaways from this project are that engineers, regulators and legal scholars define safety differently and will likely need to better collaborate to advance safe transportation systems. A blueprint for ADS Safety must be informed by experts in engineering, regulation, and law. This analysis arrived at two broad strategic ADS Safety approaches 1) reforming the Federal Motor Vehicle Safety Standards (FMVSS) or 2) the introduction of an ADS Safety Case Approach. Both approaches may include performance-based and technology-neutral standards, which are considered best practices, and both could advance safety policy.

Regarding ADS Data Policy, focusing on data collection and privacy considerations will be key to developing a workable strategy that is not thwarted by legal disputes. The analysis concludes that data collection will be critical to inform oversight and guidance to ensure safe adoption and scaling of ADS, and to advance continuous improvement of ADS industry. This data collection effort must balance the privacy of consumers and proprietary interests.

Finally, the paper highlights the importance of considering human factors and human reliability in ADS safety, particularly in the roles of safety drivers and remote operators. The analysis concludes that guidelines and requirements to protect these human operators and address safety risks associated with human-system interaction. Overall, this white paper aims to provide stakeholders across sectors a foundational understanding of ADS safety and introduces policy directions for further exploration and development in this rapidly evolving field. While this paper outlines these critical issues, there is still substantial work ahead to identify remedies and ensure safety outcomes that apply uniformly across the industry.

The following topics will advance the field:

Evaluate Domestic and International Efforts on Connected Infrastructure Needs and Develop a California Connected Vehicle Policy Framework: Connectivity may improve safety outcomes if the technology can improve communication between vehicles, infrastructure, and edge-computing devices. Connected ADSs equipped with robust and reliable technology enablers can bridge ODD gaps, removing fragmentation as vehicles navigate complex or unpredictable terrain.[187] The Australian Office of Future Transport has similarly stated support for the concept of connectivity to enhance safety at intersections and prevent collisions.[188] China is also investing in the potential of connectivity to enhance safety.[189] In the

[187] ARIA ETEMAD & PHILIPPE STEHLIK, HI-DRIVE, HTTPS://WWW.HI-DRIVE.EU/ (LAST VISITED AUG 1, 2023).
[188] AUSTRALIAN GOVERNMENT OFFICE OF FUTURE TRANSPORT TECHNOLOGY, CONNECTED AND AUTOMATED VEHICLES, HTTPS://WWW.INFRASTRUCTURE.GOV.AU/INFRASTRUCTURE-TRANSPORT-VEHICLES/TRANSPORT-STRATEGY-POLICY/OFFICE-FUTURE-TRANSPORT-TECHNOLOGY/CONNECTED-AUTOMATED-VEHICLES (LAST VISITED AUG 1, 2023).
[189] HUI ZHAO ET AL., THE DEVELOPMENT OF AUTOMATED DRIVING IN CHINA: A COMPARISON TO GERMANY REGARDING THE GOVERNMENT POLICIES, LAWS AND REGULATIONS, AND INDUSTRIES, 0 TRANSPORTATION LETTERS 1 (2023).

U.S., while some support for vehicle-to-everything (V2X) emerged, the U.S. took a possible step backwards in this area, due to recent actions by the Federal Communications Commission (FCC) to reallocate the 5.9 GHz band. Automakers are looking at other connectivity opportunities, and there is a long-running debate about the usefulness of the dedicated short-range communications versus alternatives like a broader cellular 5G network connectivity. Advances in 5G technology, and other innovations offer new options, but many high latency, low bandwidth devices are also high cost, high fault tolerance, and have slackened security.[190] 5G may present opportunities for vehicle-to-vehicle (V2V) communication as well, allowing cars to drive cooperatively and reducing risk, but ensuring the data received is authentic and there are sufficient policy structures in place will be critical to expand vehicle computing capacity. This will require investigation into the costs, challenges, and benefits of a connected vehicle network for California.

Develop a Safety Framework for ADAS Vehicles: Following on the ADS Blueprint presented in this report, it is becoming increasingly apparent that developing a Blueprint for Advanced Driver Assistance Systems (ADAS) may also be necessary. This might involve several key steps. First, research is needed to evaluate the current state of the industry by assessing the deployment and performance of various ADAS technologies, from basic lane-keeping assistance to more complex automated emergency braking. Next, a thorough analysis of the Federal Motor Vehicle Safety Standards (FMVSS) is needed to identify regulatory gaps that may not adequately address the specific risks associated with ADAS operation. This includes examining whether the standards sufficiently cover the variability in sensor technologies and system capabilities. A critical part of this process is conducting a risk analysis of different ADAS sensor technologies, such as full sensor suites (including LIDAR, radar, and ultrasonic sensors) versus camera-only systems, and analyzing the safety implications of these configurations. This will account for software advancements in AI systems, like world foundation models. Additionally, research is needed to investigate how recent AI advancements influence ADAS performance and reliability. This is crucial to understanding the safety benefits or risks introduced by these technologies. Based on these findings, enhanced safety regulations could be proposed, encompassing new testing protocols, performance benchmarks, human-ADS interaction needs, and requirements for real-time monitoring and reporting. Finally, engaging with industry stakeholders will be key —including vehicle manufacturers, technology developers, safety experts, and regulatory bodies. This is vital to build consensus on the proposed safety framework, ensuring that the regulations are practical, enforceable, and effectively enhance ADAS safety and broader public safety.

These topics will complement the ADS Blueprint in this report, however, all of this research will require ongoing efforts, as research keeps pace with the the rapidly evolving ADS, connectivity, and ADAS sectors.

[190] ID.