Rules for the Driverless Road: Developing Federal Safety Standards for Autonomous Vehicles

Introduction and summary of the issue

Autonomous vehicles (AVs) have the potential to revolutionize transportation in the United States and beyond. The benefits of AVs in creating safer, less congested roadways, as well as decreasing pollution and energy usage, have been acknowledged by academia, industry, and government officials in the last five years (Fagnant and Kockelman, 168; S. Hrg. 114-416, 1; NHTSA, "Vision," i).

There exists, however, a major barrier to the mass adoption of AVs and the realization of these benefits: a lack of public confidence in the technology. A recent report on a survey of 4,135 U.S. adults showed that while respondents had been informed about AV development (35 percent reported having seen "a lot" of coverage on AVs; 59 percent had seen "a little"), 56 percent of respondents said they did not want to ride in an AV (Pew Research Center, "Automation in Everyday Life," 29-33). Among those who said "no," 42 percent said their top reason was "don't trust it/worried about giving up control," while 30 percent said "safety concerns" (Pew, "Automation," 33).

The reasons cited above—namely, the public's perception of the risks associated with the technology and the uncertainty of outcomes in giving up vehicle control—are central to the delay in the mass adoption of AVs. Studies in risk perception have pointed to a lack of clear communication of risk information between technical experts, policymakers, and laypeople as key in creating situations where technology is perceived as unsafe for public use (Slovic et al., 92).

The absence of clear communication about AV technology's risk between industry, government, and the public is due in large part to the lack of cohesive safety standards defined and mandated by a regulatory body. While this is not the only measure that must be taken in shifting public perception about AVs, legislative action to establish well-defined standards for AV safety will at least provide the important first step of harmonizing government, industry, and public expectations of what features are necessary for the safe operation of the vehicles. Establishing standards will also help provide a legal protocol for assessing liability accidents related to the technology.

These standards must be clarified as soon as possible, because AVs are already in the operational testing phase on roadways in several U.S. states, and governments are largely playing catch-up in regulating the technology. Currently, for human-operated vehicles, driver licensing has been administered by the state, whereas enforcement has been administered by the municipality and the state. While the physical requirements of safety features are enforced at the national level, some states have additional regulations outlining acceptable or required equipment (Fagnant and Kockelman, 170). With AVs, the line between the driver and the vehicle is blurred or erased, depending on the level of automation. For the purpose of this report, the Society of Automotive Engineers' levels of automation rubric will be used, and we will focus solely on vehicles at level 5, or full automation ("Levels of Automation - SAE International Standard J3016").

In the case of fully automated AVs, an AI driving system is responsible for detecting the environment around the vehicle, analyzing these inputs, and executing physical actions by activating different parts of the vehicle. Although these systems function like the drivers whose licensing and behavior have traditionally been under the jurisdiction by state governments, it

would be impractical and unsafe to attempt to apply different rules to AI systems in different states (Swanson, 1134). While AI systems in AVs take on the active role in making and executing decisions for the vehicle and are dynamic by nature in that some use machine learning to improve their performance over time, they are ultimately a pre-coded feature of the vehicle, supplied to the user by the manufacturer. They must be regulated at the federal level, as other safety features are. Therefore, the Department of Transportation and the National Highway Traffic Safety Administration (NHTSA) must amend and extend the existing Federal Motor Vehicle Safety Standards (FMVSS) in order to establish a federal standard for AV systems (Federal Motor Vehicle Safety Standards).

With an understanding of the current state of AV technology, we articulate three standards be added to the FMVSS related to AV system safety before and after a collision event. To conceptualize these standards, we began by creating a framework for vetting their effectiveness in addressing the needs of the stakeholders in the AV space and in preserving the stated mission of the FMVSS. Using this framework, we developed three potential new standards related to crash events. While the three new standards are important examples, their development also serves as case studies for the effectiveness of our framework. A methodology for devising and enacting new standards that takes into account the needs of the stakeholders will be crucial to the long-term success of AV systems, as it can help create federal standards with concrete expectations for all stakeholders. This approach will be more effective than the current patchwork of solutions being adopted at the state level (Fagnant and Kockelman).

Based on a review of current government reports and technical literature, including NHTSA's most recent report on the topic, "Automated Driving Systems 2.0: A Vision For Safety," we find crash events to be a significant point of both interest and contention in the

legislative process (Kalra and Paddock, 2016; NHTSA, "Vision," 12-13; Canis, 2017). Research has shown that public opinion of AVs is closely tied to the moral dilemmas and potential feature malfunctions in crash scenarios (Bonnefon et al.). Correspondingly, crash preparation is also a cornerstone of the current vehicle safety policy. The FMVSS currently says that its standards are intended to protect the public "against unreasonable risk of crashes occurring as a result of the design, construction, or performance of motor vehicles" and "against unreasonable risk of death or injury in the event crashes do occur" (Federal Motor Vehicle Safety Standards). Our additions will therefore be an important step in amending the FMVSS to address AV technology while continuing its stated core mission.

The three example standards we proposed address three steps of the AV system functionality—sense, plan, act—with respect to crash scenarios. These three challenges are: (i) AV sensors' perceptive abilities are weather-dependent; (ii) current planning algorithms in AVs lack scene understanding and common-sense reasoning; (iii) AVs act non-deterministically due to machine learning and stochastic algorithms (see section 2). Standard (1): crash prevention with safety module and human-computer interface deals with challenge (i) by checking for malfunctions of various AV system components, including sensor outputs, and challenge (iii) by requiring testing of only the safety module, which behaves deterministically. Standard (2): post-crash vehicle behavior protocol also addresses challenge (iii) in that it defines the physical actions the AV system must execute in the event of a crash to protect the occupants of the vehicle. Standard (3): post-crash data retention and use protocol, along with the human-computer interface proposed in standard (1), attends to challenge (ii) by allowing human operators to see the car's reasoning of its actions both in real time and post-crash.

These standards will bring Standard 208: Occupant Crash Protection of the FMVSS—whose stated purpose is "to reduce the number of deaths of vehicle occupants, and the severity of injuries, by specifying vehicle crashworthiness requirements in terms of forces and accelerations measured on anthropomorphic dummies in test crashes, and by specifying equipment requirements for active and passive restraint systems"—into alignment with the new technical demands of AVs (FMVSS).

Many major stakeholders in the AV space—policymakers at the state and federal level, manufacturers of AV systems, and users of AV for personal and commercial purposes—are all looking to implement the systems as quickly as possible, but in a manner that limits their liability during accidents (Anderson et al., 2016). The benefits of AVs are widely touted and the companies developing AV systems are pushing for legislation that will not stifle their ability to test and innovate. Policymakers wish to position themselves as forward-thinking about AV systems, but need to enforce standards that provide a framework for liability in the event of crashes or malfunctions (NHTSA, "Vision," 20-21).

Other stakeholders, however, such as consumer protection and safety advocacy groups are pushing back against the deployment of AVs beyond the testing stage until standards are set (Rosenfield, 27). Citizens who wish to use AV systems value the technological advancement and want to get the vehicles on the road, while other citizens see AVs as unsafe and untrustworthy, and oppose their deployment until their risk assessment is proven otherwise (Pew, "Automation").

The proposed framework will allow the goal of this work to extend beyond the scope of this report, and help policymakers balance stakeholder needs in creating further AV-related amendments to the FMVSS. The proposed additions to the FMVSS are a step forward in creating

a set of federal standards that will enable the innovation prioritized by industry. Moreover, they will facilitate the widespread use of AV systems prioritized by policymakers, and satisfy the needs of citizens by meeting expectations for safety and performance, and allowing for the attribution of liability in the event of a crash.

1 Current state of AV standards and legislation

As it stands, AVs are regulated at the state level by two types of documents: legislation and regulations written by the state department of transportation interpreting the legislation and applying expert judgment. The first state to adopt AV legislation was Nevada, which in 2012 passed a measure requiring its Department of Motor Vehicles to acknowledge and develop standards for AVs. Since then, 33 state governments have passed legislation related to AVs (Weiner and Smith). Governors in Arizona, Delaware, Massachusetts, Washington, and Wisconsin have issued pertinent executive orders (Weiner and Smith).

As of December 2017, there are 31 State Senate Bills still "pending" on the issue (NCSL–National Conference of State Legislatures). However, these legislative measures vary greatly in terms of the permitted use of AVs within states—some only allow for testing, whereas others give authorization to use AVs on public roads (NCSL–National Conference of State Legislatures).

The federal government has been slower in taking legislative action on AVs. While its interest in early AV technology led to the first DARPA Grand Challenge in 2004, and it has released non-regulatory statements, notably NHTSA's "Preliminary Statement of Policy Concerning Automated Vehicles" in 2013, it has not yet moved to enact any notable legislation on AVs. Only three majors bills have been proposed in the Senate—the Autonomous Vehicle

Privacy Protection Act, introduced in November 2015 and was last referred to the Subcommittee on Highways and Transit (H.R.3876.), the Fixing America's Surface Transportation (FAST) Act enacted on December 4th, 2015, which directed the U.S. Government Accountability Office "to assess the status of autonomous transportation technology policy developed by U.S. public entities" (H.R.22.), and finally, the Safely Ensuring Lives Future Deployment and Research In Vehicle Evolution Act (or SELF DRIVE Act), which was referred to the Senate Committee on Commerce, Science, and Transportation on September 7th, 2017 (H.R.3388).

The SELF DRIVE Act states its purpose as "to memorialize the Federal role in ensuring the safety of highly automated vehicles as it relates to design, construction, and performance, by encouraging the testing and deployment of such vehicles" (H.R.3388). It seeks to amend Chapter 301 of subtitle VI of title 49, United States Code, on "updated or new motor vehicle safety standards for highly automated vehicles," in order to require the Secretary of Transportation to issue "a final rule requiring the submission of safety assessment certifications regarding how safety is being addressed by each entity developing a highly automated vehicle or an automated driving system" (H.R.3388). The rule is to include "specification of which entities are required to submit such certifications," as well as "a clear description of the relevant test results, data, and other contents required to be submitted by such entity, in order to demonstrate that such entity's vehicles are likely to maintain safety, and function as intended and contain fail safe features" (H.R.3388).

The requirements of this act contradict with the NHTSA's "Vision for Safety," which explicitly states that it is "voluntary guidance" and "encourages entities engaged in testing and deployment to publicly disclose Voluntary Safety Self-Assessments of their systems in order to demonstrate their varied approaches to achieving safety" (NHTSA, "Vision," ii). In the federal

government's limited attempts at issuing directives on AV safety standards, it is already apparent that AV manufacturers do not meet policymakers' expectations for compliance. This disconnect further makes the case that clear, cohesive legislation must be developed and enacted.

If enacted, the SELF DRIVE Act would be a step forward in establishing the federal government's jurisdiction in regulating AVs, as its third clause states a federal preemption for all AV safety legislation: "No State or political subdivision of a State may maintain, enforce, prescribe, or continue in effect any law or regulation regarding the design, construction, or performance of highly automated vehicles, automated driving systems, or components of automated driving systems unless such law or regulation is identical to a standard prescribed under this chapter" (H.R.3388).

This preemption is consistent with our recommendations for safety standards to be enforced at the federal level, but the Act does not articulate a timeline for specific action to amend the FMVSS. The Act mandates that "not later than 1 year after the date of enactment of this section, the Secretary [of Transportation] shall make available to the public and submit to the Committee on Energy and Commerce of the House of Representatives and the Committee on Commerce, Science, and Transportation of the Senate a rulemaking and safety priority plan, as necessary to accommodate the development and deployment of highly automated vehicles and to ensure the safety and security of highly automated vehicles and motor vehicles and others that will share the roads with highly automated vehicles" (H.R.3388).

Given the rapid pace of AV development in the private sector and the estimates for mass deployment ranging from 2025 to 2050, the federal government needs to respond with specific recommendations for amendments on a much shorter timeline than simply announcing a plan one year after the Act could potentially come into effect (Bansal and Kockelman, 51).

Addressing this gap, our report provides tools for developing amendments to the FMVSS in a shorter timeframe.

2 Current state of AV technology

Although AV technology has advanced substantially since its inception, with level 5 vehicles already in use in specific areas for testing purpose, they are still far from being mainstream. Among all the challenges facing AV developers, we identified three main issues, one in each step of the "sense-plan-act" design that AVs employ, similar to many other automated systems.

In a current level 5 autonomous vehicle, a suite of sensors on the vehicle gathers information about the outside world and the vehicle's relation to it. Software algorithms interpret the sensor data and plan the vehicle's actions accordingly. Finally, the vehicle's control system converts these plans into actions. Several "sense-plan-act" loops may run in parallel and at different frequencies in different parts of the car.

The most significant limitation in the sensing step is that AVs' perceptive abilities often depend on the ambient environment. The most common sensors on AVs are cameras, LiDAR (Light Detection and Ranging), and RaDAR (Radio Detection and Ranging). Cameras are the leading technology for object classification since it is the only sensor that can capture texture, color and contrast information, and high level of details (Ors, 2017). Placed on top the vehicle, LiDAR systems determine distances to obstacles located anywhere in a 360° 3D view by using laser range finders, which emit light beams and calculate the time of flight from reflection. Like LiDAR, RaDAR systems also use time of flight to determine distances to objects; however, RaDAR only works well on metallic surfaces. Complementing camera, LiDAR, and RaDAR,

ultrasonic sensors provide accurate short-range data (1-10 meters), infrared sensors allow object detection at night, and global positioning systems (GPS) facilitate vehicle localization (Anderson et al., 2016).

In the ideal situation where AV sensors have perfect perception, their planning algorithms can choose provably optimal behaviors and their execution would be much faster and more accurate than humans (Anderson et al., 2016). However, the perceptive abilities of sensors are often very dependent on weather conditions. Most camera systems are calibrated to give better performance in clear sunny days than foggy and rainy days. LiDAR accuracy can also be negatively impacted by rain, fog, and snow, which may result in unwanted detections due to backscattering from snowflakes or water droplets (Rasshofer et al., 2011). While RaDAR is not as affected by weather conditions as LiDAR, it cannot relay size and shape as accurately as LiDAR in the first place.

After AVs receive data from the sensors, the next major challenge is to make sense of these signals and plan their actions accordingly. Although current computer vision systems have impressive computing and processing power, they lack humans' sophistication in interpreting visual data. For example, a human driver knows that normally a mailbox could not be crossing a street. If she does see a mailbox-like object crossing the street, she may deduce it is actually a person in a mailbox costume. However, in extremely windy weather, a gust of strong wind could blow a mailbox across the street, and the human driver would recognize the mailbox given the context and react accordingly. Unlike humans, current artificial intelligence systems used in AVs do not possess such commonsense reasoning ability in understanding scenes, and would encounter difficulty assessing the reasonableness of the same sensor input (i.e. mailbox crossing the street) under different conditions (Gilpin, 2017).

Finally during the "act" step, in contrast to traditional vehicles that have deterministic behavior, self-driving cars may behave differently in similar situations due to machine learning and stochastic algorithms. When making decisions, AVs perform non-deterministic computations, such as a planner algorithm that ranks the results of several randomly selected candidates (Koopman and Wagner, 2016). In addition, to improve performance and reliability over time, AVs are able to learn with machine learning algorithms so that they do not make the same mistakes or errors again in the future. This unpredictability creates complications for safety verification and testing for at least two reasons: first, exercising a particular edge-case situation is difficult because one must contrive just the right conditions, often unknown to the evaluators *a priori*, that activate the edge-case behavior; second, assessing the correctness of the test results is difficult due to AVs' lack of unique behavior for a given scenario (Koopman and Wagner, 2016).

The three proposed amendments to the FMVSS standards acknowledge these challenges and intend to ensure safety despite them. The safety module in the first standard addresses the limitation in AVs' perceptive abilities by checking whether the sensor outputs are within reasonable constraints imposed by other car components that use these data. Along with the post-crash behavior defined in the second standard, it also handles the complication in testing due to AVs' non-deterministic behavior by only requiring the testing and validation of the safety module, which behaves deterministically. The data log presented in the third standard and the human-computer interface proposed in the first amendment deals with the challenge of scene understanding by allowing human operators to see the car's reasoning of its actions both in real time and post-crash.

3 The new standards as case studies for amending the FMVSS

The three standards proposed in this report serve as case studies for the application of a new framework for thinking about AV regulation. In formulating the three standards included in this report, we developed these criteria for vetting the regulations.

Is the standard...

1. Effective in preventing "unreasonable risk" in the operation of AV and in crash scenarios according to the goal of the FMVSS?
2. Specific, narrowly-tailored, and informed by in depth research into the technical components of AV operation?
3. Testable with a clear pass/fail distinction and can a regulatory authority feasibly test it at scale?
4. Practical and attainable for AV manufacturers to meet in terms of development cost, computing resources, execution, or state of the technology? Will it promote innovation in the space as opposed to being prohibitively expensive?

These questions are used to judge how effective potential standards are in satisfying the needs of the stakeholders in the AV space. In our amendments, an affirmative answer to each question was necessary for the standard to be deemed effective and comprehensive. Therefore, in developing further regulations to address other AV safety features not covered in this report, we recommend the government employ the same framework for determining if their proposed standards are optimal. Our strategy was informed by values of the stakeholders in the AV space as determined in a review of recent literature on barriers to deployment (NHTSA, "Vision;" Swanson, "Somebody Grab the Wheel!;" Anderson et al., 2016). In the sections below, we demonstrate how this framework was used to develop our three proposed standards.

3.1 New FMVSS Standard (1): crash prevention with safety module and human-computer interface

The first standard seeks to protect against "unreasonable risk" by addressing the challenges posed by the "sense" and "act" steps of AV systems. This new standard requires a built-in safety module and a user-friendly human-computer interface. The safety module checks for and responds to malfunctions of various car components, accounting for the non-deterministic behavior of self-driving cars. The human-computer interface allows operators to understand how the vehicle is interpreting the sensor data, what subsequent actions it intends to perform, and why it chooses to execute these actions.

As noted previously, AVs may behave differently in similar situations due to machine learning and stochastic algorithms. While this learning process helps AVs increase their performance and reliability in the long term, it creates challenges for deriving specific response appropriate to every scenario. To ensure safety in all scenarios without knowing the particular steps that the AV would take, we recommend a default, short duration "safing mission" when components fail or malfunction (Koopman and Wagner, 2017). This mechanism is analogous to a physical safety cage that limits the range of motion of a robot arm so that it cannot engage in harmful movements (Adler et al., 2016). An example of a "safing mission" can be auto piloting off the road with proper signals to the other cars and notifying the police. Instead of testing every component, safety certification and validation can focus on testing whether the safing mission is invoked upon component failures. Such practice could considerably reduce the overall system cost and complexity (Koopman and Wagner, 2016).

To enact the "safing mission," there should be a safety module that checks whether every component is operating properly. Two main approaches have been explored in research that

appear to be comparably feasible and acceptable: the safety "supervisor" approach and the "spy" approach.

In the safety "supervisor" approach, a simple safety supervisor monitors a set of parameter constraints and checks it against the planned behavior of the operational components (Adler et al., 2016). The supervisor itself need not be intelligent and the simplicity of its implementation can help us better establish its safety guarantees. If the planned behavior fails to satisfy the constraints, the safety supervisor detects the potential malfunction and triggers the "safing mission."

Alternatively, various car components may cross check each other (Sussman, 2017). The tradeoff for this "spy" model is that it replaces a central point of failure (the "supervisor") with a distributed system, but increases the complexity of each single module. In this approach, every module has its own expectations of how its neighboring modules should behave. When the expectations are not met, any module can trigger the "safing mission" to prevent accidents.

Although outlines of infrastructure and proofs of concept are available for both approaches, neither of them has actually been implemented in self-driving cars. According to the Society of Automotive Engineers' levels of automation rubric, level 5 AVs need to be able to handle all operational and tactical aspects of driving, even if the human driver does not respond appropriately to a request to intervene (Society of Automotive Engineers). Malfunctions of internal car components fall under this specification, so level 5 AVs must apply either one of the two proposed safety module frameworks or develop new ones to demonstrate their capability of maneuvering a component failure before being deployed on the roads.

Besides preventing themselves from crashes due to malfunctions, AVs should also be able to relate their perceptions, actions, and intentions to the human operators to demonstrate

appropriate scene understanding. Therefore, an easy-to-use human-computer interface must be implemented, so that the operators can understand the car's behavior and its reasoning for any particular action. Since level 5 AVs are expected to perform "all aspects of the dynamic driving task under all roadway and environmental conditions that can be managed by a human driver," they should demonstrate equal or stronger reasoning capacities as compared to humans so as to systematically prove its ability as a reliable full-time substitute (Society of Automotive Engineers).

Checking the vehicle's perception simply requires proper rendering of the outputs from the computer vision system. This step is already used by software developers who design and monitor the computer vision algorithm to debug and test the system. Therefore, real-time tracking of the AVs' perception is feasible and should be enforced to build human operators' trust towards the autonomous agents.

Regarding explaining its actions, the integration of a propagator network helps the car produce a causal chain of reasoning, which can be translated into human-readable explanations (Gilpin and Yuan, 2017). Gilpin and Yuan's work is the first application of qualitative reasoning for complex and comprehensive vehicle actions. Their propagator network includes a qualitative mechanical model that uses dependencies to describe relationships between different mechanical car components, and a semi-quantitative physics-based model that incorporates geometric constraints, calculates forces on wheels, and combines this information with sensor data to provide qualitative explanations.

For instance, the steering wheel is connected to a torque amplifier, which in turn is controlled by a rack and pinion mechanism. These components thus depend upon and constrain each other. The rack and pinion mechanism is linked to the front and back wheels, all of which

have normal force and friction acting on them. Measurements of these forces allow the physical model to deduce whether the car is accelerating, turning etc.

Currently, Gilpin and Yuan's models only offer post-event analysis for specific scenarios; however, with addition of expert knowledge on vehicle geometry and physics as well as optimizations on interval finding, such reasoning should be available for any vehicle actions at run-time (or more precisely, with a slight delay after the action has taken place, but this delay would not be noticeable by the human operators).

Real-time explanations for vehicle actions are crucial in providing assurance to passengers that the car perceives its environment and behaves as intended as well as in aiding the police to determine and assign faults after accidents. In a scenario where the car is suddenly decelerating, the operator may wonder why it does so. By querying this action on the interface, she may find that the car is entering a school zone with a lower speed limit. These explanations should also be retained in the car log for post-crash analysis, as detailed in new FMVSS standard (3): post-crash data retention and use protocol.

3.1.1 Evaluation of standard (1) based on the proposed framework

The three proposed standards serve as case studies for our suggested framework in thinking about AV regulation. This section presents a thorough evaluation for standard (1) by showing how it answers the four core questions in the framework. More condensed evaluation for the next two standards are included at the end of their respective sections.

1. Effective in preventing "unreasonable risk" in the operation of AV and in crash scenarios according to the goal of the FMVSS?

In line with the goals of current NHTSA standards to protect against risk of crashes and risk of injury or death when crashes do occur, this amendment is proposed to focus on achieving the

first objective. The safety module detects anomalies in the functionalities of car components before they result in crashes. Regarding the human-computer interface, although it does not directly prevent unreasonable risk, its data is kept in the car log and contributes to strengthening AV safety in the long run, as explained in section 3.3.

2. Specific, narrowly-tailored, and informed by in-depth research into the technical components of AV operation?

Standard (1) is specific as it outlines clear expectations for the safety module and the human-computer interface, along with possible ways of implementing these two components. It is also narrowly-tailored because the intention of enforcing a safety module and a corresponding "safing mission" is so that an AV does not need to exhaustively respond to edge cases separately, but can rely on one operation to handle any malfunctions. Finally, the proposed standard is informed by in-depth research, as demonstrated by the implementation suggestions drawn from literature review.

3. Testable with a clear pass/fail distinction and can the regulatory authority feasibly test this standard at scale?

The safety module is a testable component with deterministic output, namely the default "safing mission," which resolves the complication of testing non-deterministic behavior of AVs in general. While the human-computer interface is not a critical safety component, the causal-chain of reasoning recorded in the log facilitates liability assignment and safety improvement over time. To verify the readability of AV-generated explanations, the regulatory authority may administer a set of test cases based on known challenging maneuvers in regular vehicles and decide with majority opinions from human evaluators on whether they can understand the car's reasoning for its actions.

4. Practical and attainable for AV manufacturers to meet in terms of development cost, computing resources, execution, or state of the technology? Will it promote innovation in the space as opposed to being prohibitively expensive?

With the suggested implementation overviews for both the safety module and human-computer interface, this standard is certainly practical and attainable for AV manufacturers to meet. The safety module, in particular, promotes future innovation by mandating a default, deterministic "safing mission" and thus allowing other non-deterministic car behavior, so that AV manufacturers and developers can freely enhance the stochastic or machine learning algorithms to continuously improve AVs' performance and reliability.

3.2 New FMVSS Standard (2): post-crash vehicle behavior protocol

The second standard addresses challenges related to the "act" step of AV systems by defining the physical actions the AV system must execute in the event of a crash to protect the occupants.

As of today, standards for post-crash behavior in level 5 AVs are lacking, since all existing testing and reports involve vehicles with lower automation levels. Additionally, the current FMVSS standards for post-crash behavior focus exclusively on fuel integrity and flammability of interiors (FMVSS, Standard no. 301, 302, 303, 304). Full automation of level 5 AVs requires them to take more extensive ations in the event of a crash, since passengers are not expected to participate in these post-crash safety protocols. Hence, the standardization of post-crash behavior in fully automated vehicles is crucial in demonstrating their robustness in reducing unreasonable risk of injury or death of occupants, as compared to regular vehicles.

We propose an Independent Diagnostic Module (IDM), capable of sensing the functionality of individual car modules after a crash scenario, assessing the severity of the crash, and finally carrying out a deterministic response. The IDM should be impact-resistant and safely located, similar to a "black box" in a plane (Patil et al., 2013). The signals it perceives come from internal vehicle components, which allows it to constantly monitor which modules are or are not working. In this sense, IDM is similar to the safety module proposed in the first standard, and the manufacturers can choose whether one integrated module or two separate modules should be created, as long as they meet all specifications in both standards.

When a crash occurs, the IDM's planning and control systems gain precedence in the high level decision process (HLDP) of the vehicle, causing the car to react differently according to different scenarios, as detailed below. To protect passengers riding in AVs, the diagnostic

module should also present a tool capable of evaluating human health after a crash, especially human consciousness and responsiveness.

To assess post-crash scenarios and behaviors, we propose a vehicle damage scale ranging from one to five in magnitude, with increasing severity. The parameters of our scale are based on the ones published in "State of Texas Vehicle Damage Guide for Traffic Crash Investigators," modified such that they do not cover the specifics of where the damage occurred in the car, but rather gauge the overall impact of the damage to the full-scale functionality of the vehicle (State of Texas). Alternatively, parameters may be set using the Automotive Safety Integrity Level (ASIL) risk classification scheme, which we do not explore in this report.

For the rest of this section, we present the deterministic actions associated with every level of the crash severity scale. As examples, we will highlight recent crash scenarios involving AVs to show how our proposal can be applied.

The first level of the vehicle damage scale involves a light crash that does not result in any structural damage. In such case, all the sensing modules are still working correctly (i.e. no damage to cameras, RaDARs, LiDARs etc.). The IDM has simply detected a bump or unexpected scratch—for instance, a minor sideswipe, which causes minor aesthetic damage (Types of Auto Accidents). The IDM should alert the passengers of the crash detection and update the vehicle's "target location" to the closest safe pull-over zone, in order to allow the owner to assess the damage, if desired.

The first level of classification would have been applicable to the incident occurred on May 25th, 2017 involving a GM autonomous vehicle which was scratched in its read-end by a cyclist who failed to break in time for stoppage at an intersection in San Francisco (DMV Report).

The second level involves a more severe crash, which causes some structural damage to a few sensing components, but does not pose any threat to the vehicle's full functionality. An example scenario would be a low-speed rear-end collision, which impairs the sensing capability of a subset of rear cameras, without impairing the overall rear sight of the vehicle. After directing the car to stop completely, the IDM should immediately turn off the signals from the broken sensors and then assess the full functionality of the vehicle again. Subsequently, it should evaluate the passengers' health and offer assistance to call ambulance or police, if needed.

This second level would have been the case in a February 16th, 2017's incident, involving a non-autonomous Subaru crashing into a GM autonomous vehicle and causing minor damage to the GM's rear-end bumper at a San Francisco intersection (DMV Report).

Level three involves a crash with the deployment of safety airbags for the passengers. This scenario disrupts both the vehicles' sensing capabilities and its overall driving functionality. The IDM should turn off any signals coming from affected modules. Then, given the vehicle's impaired functionality, the IDM should make it stop completely, in order to minimize any other unsafe situations, and assess the passengers' health condition and offer assistance as previously described.

On February 14th, 2017, during a testing trip, Google's self-driving Lexus crashed into a bus at a speed of 2 mph. Although no injuries were reported, the vehicle's front left wheel was damaged, rendering the car possibly dangerous to drive (Wired, 2017). An IDM system, in this case, would have assigned this crash to a level three, depending on whether the car's sensing system was compromised.

For a level-four crash, the vehicle undergoes a major accident that causes most of its modules to be damaged. The IDM, however, is still working and detects all these impairments,

which induce it to immediately turn off the engine to prevent any fire hazard or other unsafe situations. If, given its working sensing capabilities, it predicts any other possibly dangerous circumstances (such as the overheating of some modules, causing possible fire hazards), the IDM should quickly alert the passengers to safely exit the vehicle. As in previous levels, it assesses the health of the passenger; however, in this case, it automatically sends the GPS location and details on the passenger's status to the police and demands assistance.

An example level-four crash occurred in March 2017, during which a Uber's self-driving Volvo landed on its side. The main focus of the IDM in this case would have been to assess the passenger's health and quickly report the details of the crash and the passenger's status to law enforcement agents.

Finally, in a level-five crash, all of the IDM's functions are disrupted and it is not able to assess anything about the car's functionality. In this case, the IDM should not override any other working module, as its lost sensing capabilities may pose more danger if given precedence over the main commands. However, if still connected, the IDM should immediately alert the police and demand assistance while assessing passengers' health, if possible. An example level-five incident dates back to May 7th, 2016, when a Tesla in autonomous pilot mode was involved in a fatal crash with a truck on a Floridian highway (Wired, 2016).

Matching this standard to our framework, the IDM is narrowly tailored to prevent "unreasonable risk" of injury or death in scenarios with different severity. It is testable at scale, similar to the safety module proposed in standard (1), and feasible both economically and technically, although manufacturers are expected to devote some extra resources to planning, building, and testing. It will promote innovation and is intended to serve as a basis for future modifications with the evolution of AV technology and the availability of more accurate data.

To ensure enforcement, the IDM should be required to store data regarding instances of crash, along with their assigned levels of damage. This information allows the federal government to verify, upon request, that different manufacturers' IDMs are working appropriately based on the specifications, and report malfunctions accordingly.

3.3 New FMVSS Standard (3): post-crash data retention and use protocol

The third standard we recommend addresses the "plan" step of AV systems. It mandates that data collected and generated by AV systems be retained in a way that allows for the appropriate parties to view and interpret the car's reasoning post-crash.

AVs generate and collect massive amounts of data, with one study putting the number around 30 gigabytes per hour (Moll et al., 2017). Some of this data must be stored within the vehicle in order for features such as the safety module proposed above to run an assessment cycle, whereas older data may be transferred to a remote server or storage facility (Bloom et al., 2017). Data ownership and privacy are points of contention in the design of AV systems, but for the purpose of this proposal, we examine only post-crash or malfunction scenarios in which data must be retained and surrendered as evidence in a legal procedure (Bloom et al., 2017).

This standard mandates the storage of any data related to a crash event, including speed of the vehicle, surrounding environment based on sensor data, functionality status of all of the car's modules and features, and output displayed to the human user via the interface leading to the crash event (NHTSA, "Vision," 14).

It also calls for a standardized federal format for the data log output, which translates the raw AV system data into a format that is readable and accurate for legal processing, similar to the crash reports used by state law enforcement. Currently, crash reporting varies between states,

with the only federal regulatory mechanism being the Model Minimum Uniform Crash Criteria (MMUCC), a set of data collection guidelines developed by NHTSA, the Federal Highway Administration (FHWA), the Federal Motor Carrier Safety Administration (FMCSA), the National Transportation Safety Board (NTSB) and the Governors Highway Safety Association (GHSA). The MMUCC guidelines identify a minimum set of motor vehicle crash data elements and their attributes "that States should consider collecting and including in their State crash data system" ("MMUCC 5th ed.").

The MMUCC includes data on environmental conditions, the number and condition of vehicles involved, the number and condition of persons involved, roadway conditions, intersection type, and actions of drivers and passengers, among many others ("MMUCC 5th ed."). The data collected by law enforcement officials at crash sites is critical in legal proceedings, particularly in assigning liability for the crash. Vehicles are often subject to a post-crash inspection by third-party investigators if insurers or drivers suspect that the crash was due to a defect in the vehicle itself.

Given that AV systems collect, store, and analyze large sets of data related to sensor inputs, AI decision-making, module functionality, and vehicle behavior outputs, if properly presented, this information could help reconstruct crash events in much greater specificity than the observations of drivers and law enforcement officials post-crash. By mandating a device in every AV system capable of aggregating and outputting this data in a standardized format, the legal process around each crash event will be informed by the same data, ensuring fairness, accuracy, and efficiency in the assignment of responsibility, which is valued by all of our stakeholders.

A set of changes to the MMUCC was proposed by NHTSA in September 2016, with one section focused on reporting crashes involving AVs. This section, however, is far from comprehensive, suggesting only the collection of the vehicle's level of automation and which automated features were engaged at the time of the crash ("MMUCC 5th ed. Second Online Forum"). Given our report focuses only on level 5 automation, the suggested changes to the MMUCC are not sufficient. Moreover, in contrast to the set of optional guidelines in the MMUCC, we advocate for mandatory regulatory measures.

Our proposed standard requires crash data to report on the AV system's "plan" leading into the crash scenario, its "plan" during the crash event, and its "plan" following the crash event, which should correspond to the execution of the post-crash behavior protocol outlined in standard (2). Here we define the system's "plan," as the linked series of decisions it makes to move from an initial state to a goal state with the capabilities of reacting to changes in a variety of factors.

Researchers have constructed a number different models for monitoring and analyzing system plans. In this standard, we use Veloso et al.'s model from the widely-cited "Rationale-based Monitoring for Planning in Dynamic Environments." Their model examines three main mechanisms characteristic of AI plans in changing conditions: monitor generation (where sensor inputs are classified as potentially relevant to the planning process), deliberation (when a change is recognized in any of the monitors, the system needs to determine whether the change should affect the plan), and plan transformations (the system may adjust the plan in several different ways: deletion of certain steps, new goals may be added or current goals modified, and prior decisions must be resolved to encompass new goals) (Veloso et al.).

These three dynamic planning mechanisms can occur many times within the span of a second, and are composed of hundreds of decisions, generating an enormous data set that must be transformed and aggregated before it is useful to a human reader. Researchers have developed some preliminary mechanisms for aggregating the data into discrete decision-making moments and presenting it as a coherent series of explanations (Molineaux & Aha; Gilpin & Yuan, 2017). These planning choices and their associated explanations could be broken into "event intervals" and stored in an event data recorder (EDR).

EDRs have been used in non-autonomous vehicles for over 40 years and in aircraft autopilot systems, with early models registering seat belt buckling, airbag deployment, warning light status and other basic parameters. The latest generation of EDRs are capable of tracking a number of pre-crash conditions, including accelerator pedal position, braking system activity, steering wheel angle, individual wheel speed, cruise control status, as well as passenger seat occupancy, driver's seat position, and diagnostic trouble codes at the time of the crash (Lange et al., 2).

AV systems have the computational power to store even more information in an EDR module. We propose the following standards to be included in an output made available post-crash by the EDR:

1. Sensor functionality: the state of all the AV's sensing mechanisms—defined as within an acceptable operational range or not within an acceptable operational range.
2. Part functionality: the state of all the vehicle's physical parts—defined as within an acceptable operational range or not within an acceptable operational range.
3. Module functionality: the state of all the vehicle's computational modules—defined by the "safing" module as within an acceptable operational range or not within an acceptable operational range.
4. 15 seconds of pre-crash event recordings (as stored in the EDR)—the time limit is taken from Lange et al.'s calculations on the appropriate EDR function for an AV system (Lange et al., 10).
5. 15 seconds of pre-crash reasoning of vehicle actions.

6. Vehicle safety feature usage: state of seat belts, doors, other non-automated passenger safety features in vehicle.
7. Road and environmental conditions: Observed both by officials at the crash scene and by the sensors inputs to the AV system (stored in the EDR).
8. All other data suggested in MMUCC should be collected by law enforcement officials.

In relation to our standard-developing framework, this standard is effective in preventing "unreasonable risk" as the retention of crash data for analysis could help identify a malfunctioning part or inform further safety features, reducing future risk of a crash. This standard is specific, narrowly-tailored, and informed by the current state of the technology as it specifies exactly which data should be retained, backed by evidence of its usefulness from the current technical literature. It is testable at scale with a clear pass/fail distinction as the regulatory body could have the system run through a simulated crash scenario and then assess whether the appropriate data were retained. Finally, while challenging and necessitating the inclusion of more secure data protection features for AVs, this standard is attainable for AV manufacturers, and will promote innovation in the space to design lean and efficient protocols for data retention and output.

Although automotive companies have traditionally pushed back against the sharing of crash data in AV systems in the interest of competitive secrecy and liability avoidance, liability assignment will actually be made more specific by this proposed standard. If the AV data log must output a description of whether modules, sensors, and parts were operating within an acceptable range pre-crash, officials will be able to determine with a new level of specificity where the malfunction arose. Since automotive companies have been working with many contractors to supply parts, sensors, and computational modules to their vehicles, this level of specificity will shift the liability from them and assign it to the respective company that manufactured the exact malfunctioned part (Boeglin, 2015).

While it is beyond the scope of this report, we understand the importance of discussing ways in which crash event data is used to improve AV system safety. There is a push for anonymized data to be disclosed to AV system manufacturers in order to inform modifications to make the systems safer, but these questions revolve around data ownership and privacy, both of which will not be directly addressed in this report (Bloom et al., 2017).

4 Addressing the argument in favor of industry self-regulation

While the FMVSS regulates the hardware safety features of regular vehicles, the industry has begun to self-regulate with regards to the development of non-AI software and electronic systems. Specifically, Automotive Safety Integrity Level (ASIL) and safety-oriented analyses outlined in the International Organization for Standardization (ISO) 26262 Road Vehicles—Functional Safety standard has been guiding the industry self-regulation (ISR). Using the automotive-specific risk-based approach for determining risk classes as defined in ASIL, ISO 26262 details the necessary safety requirements for achieving an acceptable residual risk under different classes and provides validation measures to ensure sufficient functional safety is being achieved.

Since the government is mainly playing catch-up in forming pertinent AV safety standards, ISR under ISO 26262 and future equivalent standard may be more desirable as it takes administrative and economic burdens off from the government. More importantly, it may benefit both the consumers and the manufacturers by being more sensitive to consumer needs, economical, and innovation-friendly.

First, as leaders and experts of the industry have more experience in designing, training, and testing AVs than the federal government, they are more likely to form more targeted and less

disruptive regulatory approaches that most efficiently and effectively address consumer needs (Castro, 2017). To enhance consumer confidence towards members in the industry, ISR can continually promote product quality and good commercial practices in response to consumers' expectations or complaints (OECD). In this way, the supply and demand sides of the market balance themselves to satisfy both the producers in terms of manufacturing the optimal amount and types of AVs that are feasible with the current technology, and the consumers in terms of being able to purchase AVs with the most desired features at a reasonable cost. The inefficiencies of invoking a third party to regulate the supply side are thus eliminated (Castro, 2017).

Moreover, ISR benefits the economy by creating a more flexible regulatory environment. Contrary to rigid government regulations that protect established interests, ISR guidelines often evolve over time with feedback from industry leaders and consumers, so that AV manufacturers can minimize unnecessary compliance costs (Castro, 2017). The reduced cost from the supplier side could be passed on to consumers, stimulating interest in AV usage and adoption.

Finally, bottom-up policies are more conducive to innovation than the precautionary government regulations (Thierer and Hagemann, 2014). While the latter are often preventive, requiring new innovations to be curtailed or disallowed until their developers can prove that they are risk-free, ISR permits new technologies and business models by default and thus creates less friction for innovation (Thierer and Hagemann, 2014). In concrete terms, precautionary policy making could lead to fewer choices, lower-quality goods and services, diminished economic entrepreneurialism, and slower technological progress (Thierer and Hagemann, 2014). On the other hand, ISR facilitates the deployment of AV fleets and supports the continuous improvement of AV technology.

However, ISR works well only when public and private interests are aligned, so that the regulations are not biased against one party or another (Gunningham and Rees, 1997). With investors' push to have AVs on the road as soon as possible and the immense variety of AV system features and components, it is uncertain if every AV manufacturer's definition of safety will meet the expectations of the public or prioritize their well-being (Holder, 2017). Consequently, we fear car companies may deploy AVs prematurely due to the lack of clear consequence for failing to meet certain safety thresholds, which could potentially threaten the safety of both passengers and others who share the road (Holder, 2017).

Without a third party for oversight, AV manufacturers may also lack mechanisms for review and evaluations as well as resources for improving compliance (OECD). By holding the industry more accountable with our proposed federal regulations, we prioritize public safety and better protect the values of other stakeholders—the government, the insurance companies, and the users. Acknowledging the potential advantages of ISR, we encourage the AV industry to uphold self-regulated standards in addition to the proposed federal regulatory framework. However, ISR as a stand-alone solution would not be successful in maximizing public safety or in ensuring compliance and accountability.

5 Conclusion and recommendations for future work

This report provides a meaningful contribution to the legislative framework for creating safety standards for AV systems by suggesting three specific amendments to the FMVSS related to vehicle crashes. In articulating these specific recommendations, we have illustrated a way to devise legislature for AV systems that takes into account the complex technological feasibility and the priorities of the stakeholders involved.

Documenting pre- and post-crash conditions and mandating a standard way to present these conditions so that they may be used in the legal process will allow proper risk evaluation and fair responsibility assignment. This report demonstrates that the FMVSS mission of preventing unreasonable risk to users can be adapted to AVs, and that standards will set expectations which allow for more targeted innovation within the boundaries of public and personal safety.

Much work still needs to be done in making policy recommendations around data privacy and ownership in AV systems, and a number of other standards must be added or amended in the FMVSS to cover all features of AVs. These topics are beyond the scope of this report, but our approach is applicable in developing further regulations.

We recommend that NHTSA use the four-question framework we employed to develop our three new standards in creating amendments to the FMVSS, so that they are narrowly-tailored, informed by technical expertise, testable, and innovation-friendly. When these values are prioritized, new amendments to the FMVSS will establish a well-defined standard for AV safety, promote public trust in the technology, give the industry a clear direction, and clarify the legal procedures for addressing accidents related to the technology. With the alignment of the goals and values of the stakeholders in the AV space, the vehicles will soon hit the road, and the technology's benefits will be realized in the near future.

References

Adler, R., Feth, P., Schneider, D. (2016). Safety engineering for autonomous vehicles. 46th Annual IEEE/IFIP International Conference on Dependable Systems and Networks Workshops.

Anderson, James M., Nidhi Kalra, Karlyn Stanley, Paul Sorensen, Constantine Samaras, and Tobi A. Oluwatola. (2016). *Autonomous Vehicle Technology: A Guide for Policymakers*. Santa Monica, CA: RAND Corporation. https://www.rand.org/pubs/research_reports/RR443-2.html

Bansal, Prateek, and Kara M. Kockelman. "Forecasting Americans' Long-Term Adoption of Connected and Autonomous Vehicle Technologies." Transportation Research Part A: Policy and Practice 95, no. Supplement C (January 1, 2017): 49–63. https://doi.org/10.1016/j.tra.2016.10.013.

Bergen, Mark, and Eric Newcomer. "Uber to Suspend Autonomous Tests After Arizona Accident." *Bloomberg.com*, Bloomberg, 25 Mar. 2017, www.bloomberg.com/news/articles/2017-03-25/uber-autonomous-vehicle-gets-in-accident-in-tempe-arizona

Bloom, Cara, Joshua Tan, Javed Ramjohn, and Lujo Bauer. (2017). "Self-Driving Cars and Data Collection: Privacy Perceptions of Networked Autonomous Vehicles." In *Thirteenth Symposium on Usable Privacy and Security (SOUPS 2017)*, 357–375. Santa Clara, CA: USENIX Association. https://www.usenix.org/conference/soups2017/technical-sessions/presentation/bloom.

Boeglin, Jack (2015) "The Costs of Self-Driving Cars: Reconciling Freedom and Privacy with Tort Liability in Autonomous Vehicle Regulation," *Yale Journal of Law and Technology*: Vol. 17 : Iss. 1 , Article 4. Available at: http://digitalcommons.law.yale.edu/yjolt/vol17/iss1/4

Bonnefon, Jean-François, Azim Shariff, and Iyad Rahwan. "The Social Dilemma of Autonomous Vehicles." *Science* 352, no. 6293 (June 24, 2016): 1573. https://doi.org/10.1126/science.aaf2654.

Canis, Bill. (September 19, 2017). "Issues in Autonomous Vehicle Deployment." CRS Report. https://fas.org/sgp/crs/misc/R44940.pdf.

"CA DMV–Reports of Traffic Accidents Involving an Autonomous Vehicle–OL316," n.d. https:// www.dmv.ca.gov/portal/dmv/detail/vr/autonomous/autonomousveh_ol316+.

Castro, Daniel. "Benefits and Limitations of Industry Self-Regulation for Online Behavioral Advertising." The Information Technology and Innovation Foundation (ITIF). December 2011. http://www.itif.org/files/2011-self-regulation-online-behavioral-advertising.pdf

Davies, Alex. "Google's Self-Driving Car Caused Its First Crash." *Wired*, Conde Nast, 3 June 2017, www.wired.com/2016/02/googles-self-driving-car-may-caused-first-crash/.

Dixit, Vinayak V., Sai Chand, and Divya J. Nair. (December 20, 2016). "Autonomous Vehicles: Disengagements, Accidents and Reaction Times." Edited by Jun Xu. *PLOS ONE* 11, no. 12: e0168054. https://doi.org/10.1371/journal.pone.0168054.

Fagnant, Daniel J., and Kara Kockelman. (2015). "Preparing a Nation for Autonomous Vehicles: Opportunities, Barriers and Policy Recommendations." *Transportation Research Part A: Policy and Practice* 77, no. Supplement C: 167–81. https://doi.org/https://doi.org/10.1016/j.tra.2015.04.003.

Favarò, Francesca M., Nazanin Nader, Sky O. Eurich, Michelle Tripp, and Naresh Varadaraju. (September 20, 2017). "Examining Accident Reports Involving Autonomous Vehicles in California." *PLoS ONE* 12, no. 9: 1–20. https://doi.org/10.1371/journal.pone.0184952.

Federal Motor Vehicle Safety Standards and Regulations. U.S. Department of Transportation, icsw.nhtsa.gov/cars/rules/import/FMVSS/.

Gilpin, L.H. and Yuan, B.Z. (2017). Getting up to speed on vehicle intelligence. *CSAIL Technical Report*.

Grajzl, Peter and Peter Murrell. "Allocating lawmaking powers: Self-regulation vs government regulation." *Journal of Comparative Economics* 35 (2007) 520–545. http://www.econweb.umd.edu/~murrell/articles/AllocatingLawMakingPowers.pdf

Gunningham, Neil and Joseph Rees. "Industry Self-Regulation: An Institutional Perspective." *Law & Policy*, 19: 363–414. (1997) doi:10.1111/1467-9930.t01-1-00033

Holder, Sarah. "Feds to Self-Driving Car Makers: Regulate Yourself." CityLab. Sep. 15, 2017. https://www.citylab.com/transportation/2017/09/feds-to-self-driving-car-makers-regulate-yourself/539829/

International Organization for Standardization (ISO). Standard 26262-9:201. Road vehicles -- Functional safety -- Part 9: Automotive Safety Integrity Level (ASIL)-oriented and safety-oriented analyses. Nov. 2011. https://www.iso.org/standard/51365.html

Kalra, Nidhi, and Susan M. Paddock. (2016). *Driving to Safety*. RAND Corporation. http://www.jstor.org.libproxy.mit.edu/stable/10.7249/j.ctt1btc0xw.

Koopman, P. and Wagner, M. (2017). Autonomous vehicle safety: an interdisciplinary challenge. *IEEE Intelligent Transportation Systems Magazine*, 9(1), 90-96.

Koopman, P. and Wagner, M. (2016). Challenges in autonomous vehicle testing and validation. *Procs SAE World Congress.*

Lange, Robert, Sean Kelly, Carmine Senatore, Joel Wilson, Ryan Yee, and Ryan Harrington. "Data requirements for post-crash analyses of collisions involving collision avoidance technology equipped, automated, and connected vehicles." ESV 17-0338, June 2017. https://www-esv.nhtsa.dot.gov/Proceedings/25/25ESV-000338.pdf

Molina, Caroline, Bianca Santos Tancredi. "Assuring Fully Autonomous Vehicles Safety by Design: The Autonomous Vehicle Control (AVC) Module Strategy." *IEEE Conference Publication*, 31 Aug. 2017, ieeexplore.ieee.org/document/8023692/.

Molineaux, Matthew and David W. Aha. (2015) "Continuous Explanation Generation in a Multi-Agent Domain." *Proceedings of the Third Annual Conference on Advances in Cognitive Systems.* http://www.cogsys.org/papers/ACS2015/article1.pdf

Moll, Oscar, Aaron Zalewski, Sudeep Pillai, Sam Madden, Michael Stonebraker, and Vijay Gadepally. (August 2017). "Exploring Big Volume Sensor Data with Vroom." *Proc. VLDB Endow.* 10, no. 12: 1973–1976. https://doi.org/10.14778/3137765.3137822.

"NCSL–National Conference of State Legislatures, Database on Autonomous Vehicles Legislation," n.d. http://www.ncsl.org/research/transportation/autonomous-vehicles-legislative-database.aspx

NHTSA. (September 12, 2017). "Automated Driving Systems: A Vision for Safety." U.S. Department of Transportation. https://www.nhtsa.gov/sites/nhtsa.dot.gov/files/documents/13069a-ads2.0_090617_v9a_tag.pdf.

NHTSA. (2017). "Model Minimum Uniform Crash Criteria (MMUCC) 5th Edition." U.S. Department of Transportation. https://www.nhtsa.gov/about-mmucc

Organization for Economic Cooperation and Development (OECD). "Industry Self-Regulation: Role and Use in Supporting Consumer Interests." Committee on Consumer Policy. Mar. 23,

2015.
http://www.oecd.org/officialdocuments/publicdisplaydocumentpdf/?cote=DSTI/CP(2014)4/FINAL&docLanguage=En

Patil, Chetan. "Low Cost Black Box for Cars." *IEEE Conference Publication*, 6 Mar. 2014, ieeexplore.ieee.org/document/6757114/references?part=1.

Pew Research Center, "Automation in Everyday Life." Oct 2016, Washington, D.C. http://assets.pewresearch.org/wp-content/uploads/sites/14/2017/10/03151500/PI_2017.10.04_Automation_FINAL.pdf. Accessed Nov. 26, 2017.

REPORT OF TRAFFIC ACCIDENT INVOLVING AUTONOMOUS VEHICLE. State of California, 2017. https://www.dmv.ca.gov/portal/wcm/connect/a549dbc4-a164-4813-bc72-3164261fc89a/GMCruise_021617.pdf?MOD=AJPERES

REPORT OF TRAFFIC ACCIDENT INVOLVING AUTONOMOUS VEHICLE. State of California, 2017. https://www.dmv.ca.gov/portal/wcm/connect/7775ab52-ff6a-4473-8300-e3d589cd6448/GMCruise_052517.pdf?MOD=AJPERES

Rosenfield, Harvey. "Self-Driving Vehicles: The Threat to Consumers." *Consumer Watchdog.* June 13, 2017. http://www.consumerwatchdog.org/sites/default/files/2017-10/self_driving_consumer_threat_report.pdf

Slovic, Paul, Baruch Fischhoff, and Sarah Lichtenstein. "Why Study Risk Perception?" *Risk Analysis* 2, no. 2 (June 1, 1982): 83–93. https://doi.org/10.1111/j.1539-6924.1982.tb01369.x.

Society of Automotive Engineers. (2014). "Levels of Automation - SAE International Standard J3016." SAE International. https://www.sae.org/misc/pdfs/automated_driving.pdf.

Stewart, Jack. (June 3, 2017). "After Investigating Tesla's Deadly Autopilot Crash, Feds Say Hooray For Self-Driving." *Wired*. www.wired.com/2017/01/probing-teslas-deadly-crash-feds-say-yay-self-driving/.

Stewart, Jack. (May 10, 2017). "Mapped: The Top 263 Companies Racing Toward Autonomous Cars." *Wired.* https://www.wired.com/2017/05/mapped-top-263-companies-racing-toward-autonomous-cars/

Swanson, Andrew. "Somebody Grab the Wheel!": State Autonomous Vehicle Legislation and the Road to a National Regime, 97 *Marq. L. Rev*. 1085 (2014).

Texas Department of Transportation. *VEHICLE DAMAGE GUIDE FOR TRAFFIC CRASH INVESTIGATORS*. State of Texas, 2015.

Thierer, Adam and Ryan Hagemann. "Removing Roadblocks to Intelligent Vehicles and Driverless Cars." Mercatus Working Paper, Mercatus Center at George Mason University, Arlington, VA, September 2014. http://mercatus.org/publication/removing-roadblocks-intelligent-vehicles-and-driverless-cars

Traffic Records Team, NHTSA. "MMUCC 5th Edition Second Online Forum." Governors Highway Safety Association (GHSA). http://www.ghsa.org/resources/mmucc-5th-ed

"Types of Auto Accidents | McCamy, Phillips, Tuggle & Fordham, L.L.P. | Dalton." McCamy, Phillips, Tuggle & Fordham, LLP, Motor Vehicle Accident Lawyers Serving Atlanta, Dalton, Alpharetta & Dunwoody, GA, www.mccamylaw.com/Personal-Injury-Information-Center/Types-of-Auto-Accidents.shtml.

United States. Cong. House. Transportation and Infrastructure. *Autonomous Vehicle Privacy Protection Act of 2015.* 114th Cong. 1st sess. H.R.3876. *The Library of Congress. Thomas*. https://www.congress.gov/bill/114th-congress/house-bill/3876/text

United States. Cong. House. Transportation and Infrastructure. *Fixing America's Surface Transportation (FAST) Act of 2015.* 114th Cong. 1st sess. H.R.22. *The Library of Congress. Thomas*. https://www.congress.gov/114/bills/hr22/BILLS-114hr22enr.pdf

United States. Cong. House. Energy and Commerce. *Safely Ensuring Lives Future Deployment and Research In Vehicle Evolution Act of 2017.* 115th Cong. 1st sess. H.R.3388. *The Library of Congress. Thomas*. https://www.congress.gov/bill/115th-congress/house-bill/3388/text

United States. Cong. Senate. Commerce, Science, and Transportation. *Hearing on "Hands Off: The Future of Self-Driving Cars."* Mar. 15, 2016. 114th Cong. 2st sess. (statement of John Bozzella, President and CEO, Association of Global Automakers, Inc.) S. Hrg. 114-416. Accessed via globalautomakers.org. Nov. 27, 2017.

Veloso, Manuela M., Martha E. Pollack, Michael T. Cox. (1998). "Rationale-Based Monitoring for Planning in Dynamic Environments." *Proceedings of the Fourth International Conference on Artificial Intelligence Planning Systems.* https://ocs.aaai.org/Papers/AIPS/1998/AIPS98-021.pdf

Weiner, Gabriel and Bryant Walker Smith, Automated Driving: Legislative and Regulatory Action, cyberlaw.stanford.edu/wiki/index.php/Automated_Driving:_Legislative_and_Regulatory_Action

Division of work:

Frankie: I reviewed the current regulatory measures put in place for AVs at the state and federal levels, as well as the proposed bills on AVs being circulated in Congress. I also did a search through SAE databases, as well as recently published research on crash data retention in non-AV crashes and on some efforts to adapt these protocols to AVs. I found legal papers on the challenges of assigning liability in AVs crashes and some proposed solutions to use data retention to do so. I authored the sections "Current state of AV standards and legislation" and "New FMVSS Standard (3): post-crash data retention and use protocol," and contributed to the introduction and conclusion of the paper, as well as editing the standard sections in order to create a more consistent structure in the way we presented our information. I greatly appreciated Giulio and Zoe's computer science expertise in creating my data retention standard, as it was difficult at first for me to parse the intricacies of AI planning. It was a great experience to work with them, and they were excellent teammates every step of the way.

Giulio: I mainly focused on the second standard that we proposed (FMVSS Standard (2): post-crash vehicle behavior protocol) and, partially (mainly for the first draft), on the counter argument that the industry might have with regards to our proposed regulations (self-regulated industry). I focused my research on the current state of technology as well, mainly looking at diagnostic systems in AVs, and at crashes behaviors. I found a couple of other papers that propose similar centralized systems, but not for diagnostic purposes. We together discussed the

questions of the framework, and I tried to do some editing although my editing abilities are definitely not as strong as Zoe's and Frankie's who could easily be editors at Wired.

Zoe: I focused on researching about the evolution and the current state of technology used in AVs and how it can be improved to prevent AVs from accidents. To validate the technical solutions proposed in this draft, I had extensive conversations with our mentor, Leilani Gilpin, and her faculty supervisor, Professor Gerald J. Sussman. I am the author of the following sections: "2 Current state of AV technology," "3.1 New FMVSS Standard (1): crash prevention with safety module and human-computer interface," "3.1.1 Evaluation of standard (1) based on the proposed framework," and "4 Addressing the argument in favor of industry self-regulation." I also contributed to the introduction, conclusion, and general editing of the paper.