NIST Special Publication 1900-320

Workshop Report: Consensus Safety Measurement Methodologies for Automated Driving System-Equipped Vehicles

Christopher Greer
Edward Griffor
David Wollman

This publication is available free of charge from:
https://doi.org/10.6028/NIST.SP.1900-320

CYBER-PHYSICAL SYSTEMS

National Institute of Standards and Technology
U.S. Department of Commerce

NIST Special Publication 1900-320

Workshop Report: Consensus Safety Measurement Methodologies for Automated Driving System-Equipped Vehicles

Christopher Greer
Edward Griffor
David Wollman
Smart Grid and Cyber-Physical Systems Program Office
Engineering Laboratory

This publication is available free of charge from:
https://doi.org/10.6028/NIST.SP.1900-320

September 2019

C Y B E R - P H Y S I C A L   S Y S T E M S

U.S. Department of Commerce
Wilbur L. Ross, Jr., Secretary

National Institute of Standards and Technology
Walter Copan, NIST Director and Undersecretary of Commerce for Standards and Technology

National Institute of Standards and Technology (NIST) Special Publication 1900-320

31 pages (September 2019)

SP 1900 Special Publication Series

Publications in the SP 1900 subseries present information of interest to the cyber-physical systems (CPS) community, where CPS are defined as smart systems that include engineered interacting networks of physical and computational components. The series was established in 2016 by the Smart Grid and Cyber-Physical Systems Program Office of the NIST Engineering Laboratory to provide a separate identity for CPS and Internet of Things publications, including those concerned with the foundations of CPS, CPS testbed science, and CPS applications, e.g., smart grid, smart cities and intelligent transportation. The series reports on research, guidelines, and outreach efforts in CPS, and its collaborative activities with industry, government, and academic organizations.

Disclaimer

Acknowledgement

We acknowledge the contributions to this report by the workshop participants, who represented the broad spectrum of stakeholders in the topic of automated driving safety. In particular we acknowledge the volunteers that assisting in writing and reviewing the drafts.

The writing team volunteers were responsible for taking notes during the workshop, developing the various drafts and this final workshop report document. The writing team included Sam Afsoos, Nat Beuse, Myra Blanco, Michelle Chaka, Christopher Greer, Edward Griffor, Greg Leeming, Kenneth Leonard, Tom Linn, John Maddox, Nick Royal, Omar Sheikh, Prachi Vakharia, Jack Weast, Simone Hadley Wilson and David Wollman.

Several others volunteered to serve on the 'red team' that reviewed and provided additional inputs to the draft produced by the writing team. The red team included Jackie Erickson, Khalid Halba, Aaron Jones, Shawn Kimmel, Sean Laffey, James Li, Matthew Magsamen, Ann Ngo, Aviral Shrivastava and Edward Thai.

National Institute of Standards and Technology Special Publication 1900-320
Natl. Inst. Stand. Technol. Spec. Publ. 1900-320, 31 pages (September 2019)
CODEN: NSPUE2

This publication is available free of charge from:
https://doi.org/10.6028/NIST.SP.1900-320

Table of Contents

This Page Intentionally Left Blank

1 Introduction

A workshop on Consensus Safety Measurement Methodologies for Automated Driving System (ADS)-Equipped Vehicles was held at the National Institute of Science and Technology (NIST) 25-26 June 2019 on the NIST Gaithersburg, Maryland campus, sponsored by NIST and the U.S. Department of Transportation and partners Intel, Intel Mobileye, Lyft, Ricardo, SAE International, and the Virginia Tech Transportation Institute (VTTI). It was attended by over 110 participants from the manufacturing and technology industry; safety organizations; academia; and local, state, and federal government.

The workshop consisted of keynotes, panel presentations and breakout sessions, addressing the topic from the State and Users, Developers, Standards and Testing, Methods and Frameworks, and Innovation perspectives. The scope of this workshop and this report is an assessment of the need for measurement methodologies for ADS-equipped vehicle safety. This document reports on the views expressed by the participants in these sessions and summarizes the input on the steps that can accelerate progress toward safe and trusted ADS-equipped vehicles.

1.1 Overview

The automotive industry is planning to deploy Levels 3 and above ADS (SAE J3016 - Taxonomy and Definitions for Terms Related to On-Road Motor Vehicle Automated Driving Systems) in the near term, with most major manufacturers currently targeting the early 2020s.[1] Community consensus around reliable, broadly-acceptable performance metrics for assessing ADS-equipped vehicle safety can assist in facilitating the successful achievement of these deployments. This workshop explored the need for a common method to evaluate safe operating performance of emerging ADS-equipped vehicle concepts and possible paths toward consensus. This workshop also examined concepts for safety metrics, formal models, testing protocols, taxonomies, and other approaches for measuring the safety of ADS-equipped vehicles.

ADS-equipped vehicles can, in some ways, be considered cyber-physical systems (CPS). CPS comprise interacting digital, analog, physical, and human components engineered for function through integrated physics and logic, and include smart grid, advanced manufacturing, and smart cities. In 2017, NIST released a Framework for Cyber-Physical Systems[2] that establishes the functional decomposition, the facets (groupings of activities) and the aspects (groupings of concerns) that drive the conceptualization, realization, and assurance of CPS.

[1] "The Self-Driving Car Timeline – Predictions from the Top 11 Global Automakers", May 14, 2019, Jon Walker, from https://emerj.com/ai-adoption-timelines/self-driving-car-timeline-themselves-top-11-automakers/

[2] "Framework for Cyber-Physical Systems, Volume 1, Overview" NIST Special Publication (SP) 1500-201 (https://doi.org/10.6028/NIST.SP.1500-201), "Framework for Cyber-Physical Systems, Volume 2, Working Group Reports" NIST SP 1500-202 (https://doi.org/10.6028/NIST.SP.1500-202) and "Framework for Cyber-Physical Systems, Volume 3, Timing Annex" NIST SP 1500-203 (https://doi.org/10.6028/NIST.SP.1500-203)

1.2 Background

Industry-wide, globally-accepted methods for effective safety evaluation will likely include vehicle-based measurements of both qualitative and quantitative performances. The measurement of both performances is relevant to the topic. To illustrate the difference, consider an ADS bringing a vehicle to a stop at a stop sign. The measurement of this qualitative performance is simply determining whether the vehicle comes to a stop at the stop sign or not, whereas the measurement of the quantitative performance may be the measured distance it takes to stop before the stop sign (i.e., the distance to the stop bar, or line on the road, measured from the center of the front bumper). That safety evaluation could be based on an open, consensus-based set of standards that can facilitate successful ADS deployments that interoperate safely and securely in traffic, enhance mobility, save energy, improve environmental quality, and the quality of life. The emergence of such methods can be accelerated through a community-based approach and the purpose of this workshop was to explore opportunities for such an approach.

Effective methods for ADS-equipped vehicle safety measurement can build on existing safety measurement approaches but will need to address the challenges associated with applying them to a vehicle controlled by an ADS and not a human driver. For example, participants discussed rule-based and statistical methodologies as two fundamental approaches currently used to assess vehicle safety. Rule-based assessment involves the definition of a comprehensive set of rules that precisely define what it means to function safely, and which vehicles can be empirically tested against. Statistical approaches track the performance of vehicles over millions of miles of real-world operation and calculate their probability of safe operation as an extrapolation of their observed frequency of safety violations.

Participants noted that safety assessment may be further challenged by individual technologies designed to enable ADS. An example that was cited in the discussion is deep neural networks (DNN), which are currently used to implement some ADS decision-making. DNN approximate non-linear relationships between sets of inputs and outputs; however, the statistical nature by which they do so poses challenges to verification or validation of safety. Corrections could typically be implemented by submitting them as new training data, but it is not clear how this data may impact overall performance (the new data could negatively impact performance against other previously assessed safety and non-safety related requirements). Open questions are the degree to which complete retesting to all requirements, and/or the collection of extensive, new real-world operational and simulation data be required for these updates.

The challenges enumerated above and others show that measuring ADS-equipped vehicle safety may "require a shift from traditional paradigms of vehicle safety testing to a more comprehensive view of this matter."[3] To date, academia, government, and industry have put forward a variety of initial, candidate methodologies and frameworks. These include methods for implementing safety on the system, frameworks for modeling safe driving behavior, and frameworks for measuring ADS-equipped vehicle safety. Examples of each are summarized below to frame the topics that were discussed at the workshop. The inclusion of these examples should not be construed as endorsement of them nor as a complete characterization of the candidate methodologies. It is expected that detailed information and

[3] "A Novel Method to Evaluate the Safety of Highly Automated Vehicles" Joshua L. Every, Frank Barickman, John Martin Sughosh, Rao Scott Schnelle, Bowen Weng, Paper Number 17-0076; 25th International Technical Conference on the Enhanced Safety of Vehicles (ESV) http://indexsmart.mirasmart.com/25esv/PDFfiles/25ESV-000076.pdf

examples will need to be explored further. However, the workshop provided a venue for discussion of some of the proposed methods and related activities.

Sections 1.3 through 1.6 below provide extracted summaries of four published reports. These summaries were provided to workshop registrants as background reading to facilitate discussion.

1.3 "A Novel Method to Evaluate the Safety of Highly Automated Vehicles"

Every et al. (2017)[3] introduce the concept of an Instantaneous Safety Metric and discuss the fundamental constructs of its implementation (i.e., geometric region of interest and classification of vehicle interactions).

This methodology aims to determine the action to be taken by a subject vehicle by considering the possible actions of the surrounding vehicles and the impact of those actions on the subject vehicle's choice of action. This method considers the ADS-equipped vehicle operational design domain (ODD), whether there is a choice of action for the vehicles that 'continues to provide' an escape path or a path that avoids an eventual collision and assesses the severity of the maneuver required to follow that path.

All of the above steps are taken at one point in time and the positions/headings and impact and subject vehicle choice of actions are determined for discrete points in time in the future. The space occupied by a vehicle at a given point in time, as a result of any choice of action, is referred to as the vehicle's profile. The 'possible space' of a vehicle at a future point in time is the union of all profiles of that vehicle at that point in time over all possible choices of action of the vehicle. The intersection of all profiles at a given point in time in the future is the vehicle's unavoidable space at that time.

Given a point in time in the future and two vehicles A and B, if there is no intersection between the spaces of A and B at time T this is termed an 'impossible interaction'. If the intersection is non-empty, then this is termed a 'possible interaction'. A 'critical interaction' occurs when there exists a set of profiles for traffic vehicle(s) that intersects all profiles for the subject vehicle at a given point in time and it is called an 'imminent interaction' if all possible actions by traffic vehicles result in contact with the subject vehicle at some point in the future.

There are multiple opportunities for future work associated with this approach to assessing the safety of ADS-equipped vehicle decision-making, including:

- Assessing the feasibility of the required calculations. For example, the real action sets of the vehicles involved require one to consider an infinite number of possible inputs, whether it be for lateral or longitudinal motion (i.e., braking, steering, and acceleration); and
- Extending those considered herein, full acceleration and full deceleration as well as full right and full left lateral motion, would be useful in assessing the 'worst case' scenarios associated with the actions.

One potential approach to achieving feasibility could be:

- Bounding the search for critical or imminent interactions;
- Computing the escape paths that would not require severe driving maneuvers; and
- Identifying feasible action sets that correspond to the worst cases.

Progress on relevant research questions could enable a demonstration of 'feasibility and correctness' of the approach.

The authors of this article state in their concluding remarks: "The proposed method of safety analysis provides a quantitative window into the predominantly qualitative world of subtle and nuanced traffic interactions. The authors believe that this method and approach provide a sure path toward developing a tool which can be used to quantify ADS-equipped vehicle performance in simulation, test-track, and on-road evaluations."

1.4 "On a Formal Model for ADS-equipped Vehicle Safety"

Intel Mobileye researchers proposed an open and technology-neutral model for the safety of ADS decision-making process. [4] A vehicle would be deemed to be *safe relative to the model* if its driving behaviors conformed to the model. In their paper, the authors propose a driving philosophy that states that an ADS has a *responsibility* to drive safely. The duty of the ADS is to the occupants and other vulnerable road users in the environments in which the ADS-equipped vehicle is operating. Moreover, it asserts the industry has a *responsibility* to deliver a verifiably safe product.

The proposed approach is classified as flexible. It suggests that an ADS-equipped vehicle must be *sensitive* to what it means to drive safely. Different geographies around the world may have slightly different definitions of what it means to drive safely and may have correspondingly different risk tolerances and assumed behavior for the human drivers or other agents in that region. Thus, any formal model for safety must be *sensitive* to geographical differences to ensure that ADS only make safe driving decisions as agreed to and defined by the people and governments in the different geographies in which the ADS-equipped vehicle operates.

Responsibility Sensitive Safety (RSS) is an open and transparent technology-neutral, formally-verifiable model for safe decision making of an ADS that features specific parameters that can be adjusted and tuned as desired by the people and governments in which ADS-equipped vehicles may be deployed.

RSS is based on human centered concepts of what it means to drive safely. These concepts include: always maintain a safe following distance; be cautious in areas where other vehicles or pedestrians may be included; and right of way is given, not taken.

Although these common-sense human principles for safe driving are not found in any legal 'rules of the road,' the American Association of Motor Vehicle Administrators (AAMVA) has documented many of these and they generally represent a set of safety principles that are widely understood and accepted by humans as the safe way to drive. It is suggested in the paper that humans following these principles are considered as safe drivers today. Conversely, humans that follow a lead vehicle too closely, are careless in areas where others may be present, or aggressively take the right-of-way from others are considered as not driving safely, and as often the cause of crashes. This paper defines the following concepts relative to the RSS model:

- **Driving Safely**: A vehicle is driving safely if its decisions conform with the RSS model.
- **Driving Unsafely**: A vehicle is driving unsafely if its decisions violate the RSS model.
- **Dangerous Situation**: A state of the vehicle such that there is the possibility of an accident and is typically the result of some unexpected behavior from other road users.

[4] "On a Formal Model of Safe and Scalable Self-driving Cars" 27 Oct 2018 version, Shai Shalev-Shwartz, Shaked Shammah, Amnon Shusha., Cornell University Library, https://arxiv.org/pdf/1708.06374.pdf

- **Safe State**: A Safe State is a measure of the current state of the ADS-equipped vehicle with respect to all other relevant agents around the vehicle such that no Dangerous Situation exists. An ADS-equipped vehicle in a Safe State by definition can neither be causing nor involved in a Dangerous Situation.
- **Danger Threshold**: The Danger Threshold is the moment in time at which a vehicle transitions from a Safe State into a Dangerous Situation. In other words, the danger threshold is the moment in time at which a vehicle began driving unsafely by initiating a Dangerous Situation.
- **Proper Response**: A Proper Response is what the ADS-equipped vehicle could do to get out of a Dangerous Situation and return to a Safe State.

The purpose of publishing the model openly, in a series of academic papers, is not to claim that the model is perfect. Instead, the goal is to help start an open/honest conversation about safety verification of ADS decision-making across the community so that we can arrive together at a common solution for verifiably safe vehicles. Research is needed to compare the model against new scenarios, to determine whether RSS's "Safe State" can be realized under all ODDs and whether a complete RSS rule set can be specified or how this rule set can be fully implemented and validated in the real world.

The RSS model emphasizes safe driving principles. Additional topics for future research along these lines would include:

- What could be our method for developing a set of safe driving principles?
- How are we to validate these principles?
- How are we to test whether an ADS-equipped vehicle adheres to the principles?

1.5 "Measuring Automated Vehicle Safety, Forging a Framework"

A 2018 RAND Corporation report[5] defines a framework for measuring ADS-equipped vehicle safety. It identifies where measurement in simulation, closed-course, and public roads testing is possible and/or meaningful for safety, and it characterizes measurement options in terms of their feasibility, reliability, validity, and non-manipulability. It is stated that "This report is intended to foster broader understanding and discussion of ADS-equipped vehicle safety that can aid the public and policymakers in the debate over a new product category expected to reshape economic, social, and community activity."

Summarizing all of the RAND team's findings and their proposed framework is beyond the scope of this document. Instead, it is more instructive to summarize the pertinent recommendations made in this report to the community working with ADS development and testing. In addition to the proposed framework for technology and vendor-neutral safety measurement, other recommendations are:

1. Develop a formal protocol for the demonstration of ADS (in simulators, on private test tracks and on public roads with safety drivers) to facilitate comparisons across companies and evidence of safety for the public and policymakers.
2. During development, demonstration, and early deployment, when sufficient exposure has not been accumulated to allow for statistically valid comparison of rates, outcomes (e.g., crashes or an absence of crashes) could be evaluated as case studies.
3. Enable data-sharing (both between companies and with government) to improve safety across the industry and to enable real-time sharing (e.g., around roadworks). Regarding outcome

[5] "Measuring Automated Vehicle Safety, Forging a Framework" Laura Fradde-Blanar, Marjory S. Blumenthal, James M. Anderson, Nidhi Kara, RAND Corporation, 2018, ISBN: 978-1-9774-0164-9 https://www.rand.org/content/dam/rand/pubs/research_reports/RR2600/RR2662/RAND_RR2662.pdf

measures, a protocol for reporting to government entities could be codified in terms of measures, context, format, frequency, data security, governance, and other factors.
4. A formal taxonomy around ODD is needed. Such a taxonomy could specify how ODDs convey where, when, and under what circumstances the ADS-equipped vehicle can operate.
5. Given the challenges of measuring safety where the system changes, constantly and at irregular intervals, more research is needed.

1.6 "A Framework for Automated Driving System Testable Cases and Scenarios"

The following summary is extracted from a 2018 NHTSA report.[6] The goal of this research was to develop an example of a preliminary test framework for ADS that are in development and may come to market in the near to mid future. The following steps were conducted to support the development of the sample test framework.

1. Identify concept ADS;
2. Identify attributes that define the operational design domain (ODD);
3. Identify object and event detection and response (OEDR) capabilities; and
4. Identify and assess failure modes and failure mitigation strategies.

Technologies of interest in this work included light-duty automated driving functions that fell within Level 3 (L3) to Level 5 (L5) of the SAE levels of driving automation (SAE International, 2018[7]). The functions were identified based on prototype vehicles and conceptual systems. A literature review which included popular media, press releases, technical journals, and conference proceedings was performed. This review identified potential concept ADS being developed or proposed by original equipment manufacturers (OEMs), suppliers, technology companies, and other organizations. The identified ADS were categorized into a set of generic names. The terminology was modified to ADS features (as opposed to functions) to more closely align with the standardization community's language. Twenty-four conceptual features were identified, and although a thorough search was conducted, the list is not exhaustive.

The ODD describes the specific operating domains in which the ADS is designed to function. The ODD may likely vary for each ADS feature on a vehicle and specifies the condition in which that feature is intended and able to operate with respect to roadway types, speed range, lighting conditions, weather conditions, and other operational constraints. The ODD is specified by the technology developer, and the ADS could be able to identify whether it is operating within or outside of that ODD. This taxonomy is hierarchical and includes the following top-level categories:

1. Physical Infrastructure;
2. Operational Constraints;
3. Objects;

[6] "A framework for automated driving system testable cases and scenarios" (Report No. DOT HS 812 623). Thorn, E., Kimmel, S., and Chaka, M. (September 2018). Washington, DC: National Highway Traffic Safety Administration. https://www.nhtsa.gov/sites/nhtsa.dot.gov/files/documents/13882-automateddrivingsystems_092618_v1a_tag.pdf

[7] "Taxonomy and Definitions for Terms Related to Driving Automation Systems for On-Road Motor Vehicles" J3016_201806. https://www.sae.org/standards/content/j3016_201806/

4. Connectivity;
5. Environmental Conditions; and
6. Zones.

Object and event detection and response (OEDR) refers to the subtasks of the dynamic driving task (DDT) that include monitoring the driving environment (detecting, recognizing, and classifying objects and events and preparing to respond as needed) and executing an appropriate response to such objects and events, i.e., as needed to complete the DDT and/or DDT fallback.[8]

Potential failures could have significant impacts, ultimately resulting in collisions that could damage the vehicle or harm its occupants or other roadway users. Failure mitigation strategies, including both fail-operational (FO) and fail-safe (FS) techniques, were then identified and analyzed. FS techniques are used when the ADS cannot continue to function and may include options such as the following: transitioning control to fallback-ready user; safely stopping in lane; and safely moving out of travel lane/park. FO techniques can be used to allow the ADS to function at a reduced capacity, potentially for a brief period of time or with reduced capabilities, and may include options such as adaptive compensation – weighting data from a complementary component or subsystem more heavily (e.g., weighting camera data more heavily if lidar fails). The appropriate failure mitigation strategy is highly dependent on the nature of the failure and the initial conditions under which the failure occurs.

2 Workshop Description

The workshop's goal was to explore development of methodology for ADS safety that would be consensus-based and technology-neutral. ADS safety is widely seen as a major challenge, requiring broad collaboration and support. A government-industry partnership sponsored the workshop, bringing together stakeholders from across the ADS-equipped vehicle community to discuss and to articulate next steps to address this challenge. These stakeholders included participants from manufacturing, university research, transportation services, the insurance industry, testing sector, standards setting organizations, federal and state governments, safety and professional organizations, and more. A webcast of the workshop's plenary sessions was also provided and is available through the workshop website.[8] The workshop's panel sessions focused on current and projected states of ADS-equipped vehicle safety assessment methodologies. The panel sessions also addressed the complexity of ADS-equipped vehicle safety. In breakout sessions, workshop participants provided their individual perspectives on a variety of questions relevant to ADS-equipped vehicle safety, including definitions, measurement methods and metrics, best practices, role of humans in ADS-equipped vehicle safety, and possible next steps. The results of the breakout sessions were reported in plenary sessions on both days of the workshop. This section of the workshop report provides a summary of the plenary and panel sessions.

2.1 Panel: State and Users Perspectives

(Greg Leeming, moderator)
Gummada Murthy — American Association of State Highway and Transportation Officials (AASHTO)
Kevin Biesty — Arizona Department of Transportation
Marisa Walker — Arizona Commerce Authority
Jessica Cicchino — Insurance Institute for Highway Safety

This panel session aimed to characterize how state governments, municipalities, and the insurance industry are approaching the deployment of ADS-equipped vehicles. Arizona state officials described their evolution from "no standards" in 2012, to those now governing ongoing testing and use of ADS-equipped vehicles. Initially, the legislature and governor supported state agencies to collaborate with the ADS-equipped vehicle industry to determine needs and gain an understanding of the technology. Subsequent legislation and executive orders provided directions for ADS-equipped vehicle testing, including first responder protocols, resulting in active engagement of municipalities. State officials reported that state and municipalities' policies are aligned on ADS-equipped vehicles. Today, ADS-equipped vehicle companies are testing vehicles for delivery and truck platooning. Arizona's governor also established the Institute for Automated Mobility,[9] enabling government, industry and universities to address safety. The institute also facilitates public and private partnerships for developing commercial potential. Presently, the insurance industry funds the Insurance Institute for Highway Safety[10] to research ADS safety. The IIHS is currently developing a consumer rating system for SAE level 2 driving

[8] https://www.nist.gov/news-events/events/2019/06/consensus-safety-measurement-methodologies-ads-equipped-vehicles

[9] https://www.azcommerce.com/iam/

[10] https://www.iihs.org

automation features. An active forum for collaboration and information is the Cooperative Automated Transportation (CAT) Coalition[11] that serves as a collaborative focal point for federal, state, and local government officials; academia; industry; and their related associations to address critical program and technical issues associated with the nationwide deployment of connected and automated vehicles on streets and highways.

2.2 Panel: Developers Perspective — Manufacturers and Technology Companies

(Ed Griffor, moderator)
Jack Weast — Intel Mobileye
John Maddox — Lyft
Ron Medford — Waymo
Colm Boran — Ford Motor Company
Steve Kenner — Uber
Padma Sundaram — General Motors
Nick Royal — Ricardo Innovation

In this panel session, industry representatives addressed approaches toward ADS safety and cybersecurity. Multiple approaches to ADS safety were described, including scenario-based, ethics-based, mileage-based, and more. Multiple panelists acknowledged that current International Standards Organization (ISO) standards, ISO 26262 (Road Vehicles – Functional Safety) and ISO/Publicly Available Specification (PAS) 21448 (Road Vehicles – Safety of the Intended Functionality) may not be sufficient for ADS. Some panelists identified Underwriters Laboratory 4600 (Standard for Safety for the Evaluation of Autonomous Products) as a method that proposes to standardize the content in a safety assessment report using a safety case approach. Panelists also described general needs regarding ADS safety:

- Defining what it means to drive safely and associating the definitions with a methodology to assess performance. Such a definition should be formulated collaboratively.
- Determining metrics may help lead to ways to demonstrate safety. Such metrics should be standardized and repeatable, and apply to physical testing and simulation, in all development phases. Such metrics should be meaningful to the engineer and the public.
- Incorporating rules of the road into a safety program and framework. While rules vary with jurisdictions, they also have overlaps.
- Developing a safety framework that is congruent with public expectations, including those regarding off-road safety of pedestrians and cyclists.
- Determining acceptable and achievable levels of safety for early ADS-equipped vehicle deployment, thus allowing realization of real-world safety lessons after deployment.
- Discovering unintended consequences design approaches to prevent or mitigate them.

A panelist provided information that, in the United Kingdom, industry is pursuing a consumer rating for vehicles' cyber resilience. This approach has involved identifying good practices across the lifecycle of the vehicle and communicating them to the public.

[11] https://transportationops.org/CATCoalition

2.3 Panel: Current US and International Standards and Testing Methods

(Myra Blanco, moderator)
Edward Straub — Automated Vehicles Safety Consortium, SAE International
Myra Blanco — Virginia Tech Transportation Institute (VTTI)
Aviral Srivastava — Arizona State University
Brian Williams — Massachusetts Institute of Technology

This panel session provided information on existing safety standards for ADS-equipped vehicles, as well as ongoing initiatives for future safety assessments. The following SAE standards addressing existing ADS-equipped vehicle safety approaches were described:

- J3016: Taxonomy and Definitions for Terms Related to Driving Automation Systems for On-Road Motor Vehicles
- J3131: Automated Driving Reference Architecture
- J3164: Taxonomy and Definitions for Terms Related to Automated Driving System Behaviors and Maneuvers for On-Road Motor Vehicles

A future SAE Standard may address cooperative automation, which merges aspects of ADS and communication. SAE recently established the Automated Vehicle Safety Consortium, bringing together industry partners to advance safety testing for ADS-equipped vehicles. NHTSA has supported VTTI work to identify barriers to building Level 4 and 5 ADS-equipped vehicles – specifically vehicles without manually operated driving controls, or “ADS-dedicated vehicles.” The goal of this research is to develop technical translation options of the Federal Motor Vehicle Safety Standards (FMVSS) and related compliance test procedures for ADS-equipped vehicles. The project is also identifying regulatory barriers to self-certification and compliance verification of innovative new vehicle designs equipped with ADS. In terms of future testing, researchers at Arizona State University propose programmable test tracks, which could create conditions for evaluating ADS-equipped vehicle in varying environmental conditions — rain, snow, etc. – and at key intersections. MIT researchers are developing a risk assessment program for ADS-equipped vehicles.

2.4 Panel: Evolving Methods and Frameworks

(Jack Weast, moderator)
Frank Barickman — National Highway Traffic Safety Administration (NHTSA)
Jack Weast — Intel Mobileye
Michelle Chaka — VTTI

This panel session covered some of the ongoing developments of frameworks and methods for assessing the safety of ADS-equipped vehicles. For example, NHTSA has been conducting research that considers the possibility of using an "instantaneous safety measure" (ISM) to define a metric for the safety of an ADS. It is the measurement of performance, not related to the optimum situation, which has a probability of collision occurring, and actions that could be taken. Basically, this initiative considers four possible vehicle actions in a dangerous situation: brake, accelerate, steer right or steer left. These spaces where the vehicle could end up are described as the "region of possibility." ISM is calculated using a model, which also considers the possibility of collision with another vehicle or object. This region of possibility is seen as a metric. Intel Mobileye has developed a safety model, called RSS — Responsibility-Sensitive Safety — that is a formally-verifiable model for safe decision making by ADS-equipped vehicles based on common sense rules of the road. VTTI is leading the Automated Mobility Partnership (AMP) program. This program includes thirteen companies and aims to promote the development of tools, techniques, and data resources to support the development and evaluation of ADS features. It is leveraging Nationalistic Driving Data (NDD) to develop a Scenario Library of crashes, near-crashes and epochs (ADS relevant driving) cases. Analytics are being used to inform case-based decision making with rates of events and descriptive kinematic analysis. Additionally, virtual tools are being used to reconstruct the cases to support the development and evaluation of ADS performance. VTTI's AMP initiative is developing building blocks that could be used to support the development of a safety methodology.

2.5 Panel: Collaboration for Innovation

(Ken Leonard, moderator)
Finch Fulton — U.S. Department of Transportation (USDOT)
Dee Williams — NHTSA
Debbie Hersman — Waymo

This panel session addressed the role as well as importance of collaboration in ADS-equipped vehicle safety development. USDOT's approach to ADS-equipped vehicle safety is addressed in its *Preparing for the Future of Transportation (Automated Vehicles 3.0).*[12] Its goal is the integration and safe operation of ADS-equipped vehicles on U.S. roadways. Achieving this goal requires continuing collaboration with ADS-equipped vehicle stakeholders. *Automated Vehicles 3.0* seeks a data exchange with stakeholders, enabling ADS-equipped vehicle safety, as well as, improved general safety on roadways. Collaborating with stakeholders, USDOT reached an agreement on areas and formats for data exchanges (e.g., work zone activity data initiative). USDOT's plan for ADS-equipped vehicle integration onto roadways may be a general path forward, also based on stakeholder collaboration. In the meantime, USDOT encourages companies to collaborate with it in their disclosure of voluntary safety self-assessments (VSSA), based on existing industry consensus standards (e.g., SAE, ISO, etc.). USDOT provides a flexible, non-prescriptive template and examples of specific underlying science (i.e., current road safety risk and safety industry consensus standards) for those assessments. Within industry, there is growing consensus regarding the need to widely collaborate on a safety framework for ADS-equipped vehicles. Previously, the automotive, aviation, and pipeline industries used broad collaboration to overcome major challenges. Industry recognizes that such collaboration can accelerate work toward useful ADS-equipped vehicle safety assessments.

[12] "Preparing for the Future of Transportation: Automated Vehicles 3.0" U.S. Department of Transportation https://www.transportation.gov/av/3/preparing-future-transportation-automated-vehicles-3

3 Workshop Discussion Themes

3.1 What are appropriate definitions of 'safety' in a measurement context, including whether it may be a system measure, a component measure (hardware, software, etc.), a behavior/performance measure, or some combination of these?

Many stakeholders seemed to agree that an appropriate definition of safety for an ADS-equipped vehicle is comprised of a multitude of factors. Vehicle safety has multiple levels and a broad range of areas, including crash avoidance, driving behaviors, personal safety (cabin safety without a driver), occupant protection, occupant psychological effects, and cybersecurity. Each of these safety aspects were discussed at the workshop and may need to be addressed by ADS-equipped vehicle developers. During the discussion, some attendees noted that the focus for defining safety in a measurement context may need to be scoped. Throughout the workshop, there were three levels of safety discussed: (1) the engineering level, (2) the level of general public acceptance, and (3) the level of risk (e.g., foreseeable and unforeseeable, preventable and unpreventable). A short definition was proposed by some workshop attendees as follows: Safety is the avoidance of unreasonable risk in a predictable manner (based on ISO 26262). However, there was no consensus on such a definition or ISO 26262 as a starting point. As part of the scoping exercise for this question, it was suggested that the effort be focused on driving automation system Levels 4 and 5. Moreover, participants suggested that Level 3 ADS-equipped vehicles might need different considerations due to the need for a fallback-ready user.

3.1.1 SYSTEM PERFORMANCE AND COMPONENT CONSIDERATIONS

Many of the attendees expressed the view that, regardless of the metrics, the measures used should be at the system performance level rather than at the component level. From an engineering perspective, attendees stated metrics are needed to calculate whether ADS-equipped vehicle performance is within an acceptable level of safety.

Participants cited several activities and efforts for evaluating component measurements that have already been developed or for which development is on-going. System safety and system performance requirements could be the key area of focus for consensus safety measurements. System level safety may also capture aspects of component design. Currently, system safety captures how a vehicle is designed and built and not necessarily its system behaviors. Measurements of safety could support assessment of system deployment readiness without pointing to specific component standards or requirements (i.e., system- level vs. components-level evaluation). Therefore, component standards such as ISO 26262 could evolve over time to accommodate ADS-equipped vehicle behaviors and assist with the identification of risks for certain hazards. It was acknowledged that component measurement is necessary but not sufficient; ISO 21448 is also under review to address this issue, but both standards would take substantial time to revise. UL 4600 is being developed to be a standard that can evolve as technology advances and may support relatively rapid revisions.

Some attendees noted that there are hazards that present a continuum of potential risks. Accordingly, they stated that the ways in which system design might address a portion of criticality given the outcome may not be binary. The range or degree to which risk is accepted as reasonable is a key aspect of defining safety; for example, different countries or jurisdictions may identify different levels of acceptable risk for ADS. Potentially, the situations of interest may include crashworthiness and pedestrian and vulnerable road user protection given that crash avoidance might not be feasible. Aspects such as unintended consequences and the probability of exposure and/or failure were also mentioned as some of the potential considerations for defining a metric to assess the safety of an of ADS-equipped vehicle.

Different ADS-equipped vehicles are likely to have varying Operational Design Domains (ODDs). Attendees suggested that testing and measurements could consider the relevant ODDs. The NHTSA's Testable Use Cases project considered domains and appropriate behaviors.[13] Further,[14] some workshop participants noted that additional work would be needed to look at a distribution of test cases and that the challenge with scenario-based testing is that ADS-equipped vehicle developers may be inclined to design to the test. This issue could be addressed through test variation (e.g., varying certain parameters). An industry standard often defines minimal performance — while for market competitiveness, ADS-equipped vehicle developers may have additional internal requirements. Other participants noted that designing for a set of tests representing the potential scenarios for a generic domain is not a concern. For example, today's manufacturers do not design occupant protection for every crash scenario. However, the basic crash tests they design represent most of the crash situations that a vehicle could encounter. Aspects such as high severity-low occurrence versus low severity-high occurrence are considered today and could likely be considered as part of future frameworks.

3.1.2 COMPARISON TO HUMAN DRIVER

One of the goals of ADS-equipped vehicles is to advance motor vehicle safety and move closer to a zero-fatality vision. Today, the majority of crashes are attributed to human driver error.[15] One aspect of measuring safety identified by participants may be to compare the ADS-equipped vehicle's behavior to the behavior of a human driver. Are ADS-equipped vehicles equivalent to or safer than human drivers and/or perhaps other ADS-equipped vehicles? Comparing current driving population performance with new technology implementations is something that NHTSA, IIHS, and other safety organizations have undertaken for many years. However, participants identified challenges that need to be addressed when assessing ADS-equipped vehicle safety, such as what database (e.g., an ADS-equipped vehicle developer's data) or which type of human driver (e.g., teenage, drunk, older, professional, region/country differentiators) is used for comparison. As was noted at the conference, human driving consists of a complex set of circumstances, interactions and variables with respect to: the human driver (e.g., alert, fatigued, impaired, distracted, inexperienced, varying control input, headway selections, general aggressiveness), the roadway environment (e.g., weather, time of day, signage, roadway

---

[13] "A framework for automated driving system testable cases and scenarios" (Report No. DOT HS 812 623). Thorn, E., Kimmel, S., and Chaka, M. (September 2018). Washington, DC: National Highway Traffic Safety Administration. https://www.nhtsa.gov/...dot.../13882-automateddrivingsystems_092618_v1a_tag.pdf

[14] "A framework for automated driving system testable cases and scenarios" (Report No. DOT HS 812 623). Thorn, E., Kimmel, S., and Chaka, M. (September 2018). Washington, DC: National Highway Traffic Safety Administration. https://www.nhtsa.gov/...dot.../13882-automateddrivingsystems_092618_v1a_tag.pdf

[15] "Critical Reasons for Crashes Investigated in the National Motor Vehicle Crash Causation Survey" in Traffic Safety Facts, Hatioanl Highway Traffic Safety Administration (February 2015) available at https://crashstats.nhtsa.dot.gov/Api/Public/ViewPublication/812115

geometry, other road-way users), and vehicle design (e.g., different human-machine interfaces, visibility, advanced driver assistance systems - ADAS).

Participants stated that the creation of generic human comparison models could be complex. Would the required level of unavoidable risk be established as being as good as a [very good] human or as being better than [an average] human? Some workshop attendees noted that currently acceptable human performance may not be acceptable for a system that is pre-programmed to act in humans' best interest. However, there was no consensus about what level of good driving behavior was good enough. Moreover, there might be safety critical events that involve an ADS-equipped vehicle in which the event is attributable to the behavior of a legacy vehicle driver (e.g., legacy vehicle rear-ends an ADS-equipped vehicle). In this regard, workshop participants discussed how ADS-equipped vehicles interact with other road users generally, including other human drivers and public safety officials (law enforcement, fire, and emergency medical services (EMS)).

The second level of safety discussed was public perception. How the general public views safety may be based on their perceived risk. For example, if the ADS-equipped vehicle gives the impression of a potential crash by braking perceived by a passenger as abrupt and stopping at a distance perceived as too close to another object, the ADS-equipped vehicle might be judged by the passenger to be unsafe when from the ADS perspective there may have been no chance of a crash. Workshop attendees considered whether humans would accept machine error or if an ADS-equipped vehicle would have to perform flawlessly. Most human drivers believe themselves to be good drivers and would not want to be exposed to additional risk.

Some attendees believed that the development of safety measurements in advance of deployment may be needed to help inspire public confidence. Attendees believed that congruency between public expectations (e.g., do not crash, do not cause others to crash, protect occupants, obey traffic laws, get where you want to go) and real world performance may be important and the measurement for safety could be defined with this aspect in mind. These attendees believed that, if ADS-equipped vehicles are not perceived as safe, the public might be less willing to use ADS-equipped vehicles. Attendees suggested subjective metrics, such as public perception, could be objectively measured in a more comprehensive evaluation, where a deceleration profile suitable for the human user is developed and used as a comparison. A word of caution was provided by several attendees given the subjective nature of some metrics. For example, a deceleration profile might be perceived as safe for a near-crash situation but not for stopping at a stop sign. It was suggested that the ADS-equipped vehicle could detect when the environment is changing and adapt its behavior to that environment. Attendees suggested establishing a more objective way of assessing safety and staying away from aspects that might drift into comfort metrics (e.g., quality of the ride). However, it was noted that the notion of unreasonable risk is partially based on social norms, culture, and driving environment situations, which may be subjective at times.

3.1.3 REGIONAL VS. INTERNATIONAL SAFETY CONSIDERATIONS

Various participants expressed the belief that appropriate definitions of safety in a measurement context could be influenced by geographic location. State and regional safety variations already exist. State laws related to motor vehicle safety vary by state, for example, ranging from differences in enforcement of seatbelts and motorcycle helmets to speed limits and hazard flasher activation, and also variations in traffic rules (e.g., some states allow U-Turns unless otherwise noted, while others disallow U-Turns unless explicitly noted). In addition to differences at the local, state, and regional level, safety definitions and regulatory approaches may also vary internationally. For example, Europe operates under a type-approval regulatory regime, which requires that vehicles be pre-approved, while the U.S. relies on self-certification. Participants stated that considerations for how safety may be measured and viewed internationally might be important to ensure future harmonization.

The United Nations Economic Commission for Europe (UNECE) World Forum for Harmonization of Vehicle Regulations (WP.29) provides a medium for this harmonization. It allows contracting parties to participate in regulatory sessions focused on the development of regulatory standards concerning motor vehicles and motor vehicle equipment. The U.S. is a contracting party to the 1998 agreement, in which NHTSA participates (for example, see footnote below[16]).

Some participants stated that, once there is a U.S. community consensus on defining safety in a measurement context, it may then be important to share these ideas globally through forums such as the WP.29. Further, it was stated that a community consensus may allow a more rapid development and implementation of standards and associated metrics.

3.1.4 MEASUREMENT SPECIFICS

Some participants stated that the approach for metric development could help ensure safety while not impeding ADS-equipped vehicle deployment and could allow for the technology to evolve. During the workshop, there were several proposed approaches and ideas presented for developing a safety assessment method. These include the NHTSA's Instantaneous Safety Metric (ISM), Intel's Responsibility Sensitive Safety (RSS), NHTSA's Testable Use Cases project, and VTTI's Automated Mobility Partnership (AMP). Participants from the workshop expressed that the approaches had merit and could be investigated further to understand how these efforts could inform the definition of a safety measurement. It was observed by workshop participants that safety measurement for ADS-equipped vehicles shares similar challenges as security and privacy, and that safety hazards, security threats, and privacy intrusions will continue to evolve. These areas typically use risk-based models as a measurement for assessing safety. The application of procedures to mitigate security risk to people and property could be leveraged to develop some acceptable risk models for a given use case or scenario. According to some participants, risk-based modeling would not be new to NHTSA or the industry, as these methods are already used in establishing measurement criteria for crash dummy development. The application of probabilistic mathematics to assess system safety is a possible approach for defining safety and assigning risk to potential ADS-equipped vehicle behaviors at a systems level. The frequency and severity of real-world incidents may be different than the calculated risk. This may be due to the difficulty of determining the probability of extremely rare, high consequence events, for example. These events often lack an empirical basis for evaluation. As a result, it is likely that this approach will work for common risks that happen frequently, but significantly underestimate risk for rare events. Participants also noted that the means by which probability is calculated is a key challenge and may require further investigation. The probability calculation could include aspects such as severity and other environmental elements. Currently, there are frameworks that describe a crash and the elements that can be used to evaluate them (e.g., roadway's level of service). Participants noted that ADS-equipped vehicle safety measurements could take similar elements into consideration.

[16] https://www.unece.org/fileadmin/DAM/trans/doc/2019/wp29/WP29-177-19e.pdf

3.1.5 ADDITIONAL REMARKS

The following are some additional remarks presented by the workshop attendees:

- Vulnerable road users, such as pedestrians, are an important aspect in motor vehicle safety that should be considered in metrics development.
- How an ADS-equipped vehicle influences other legacy vehicles is an aspect that should be considered to avoid unintended consequences.
- Outreach and education are important to ensure that safety expectations are calibrated between users and developers. Marketing and consumer education techniques could ensure that ADS-equipped vehicle information is provided clearly and correctly to future ADS-equipped vehicle users and other road users who might share the road with these vehicles. The ADS should also understand the responsibility of other players (e.g., cyclist, pedestrians).
- Aspects of a near crash may be considered as a part of a safety metric.[17]
- There may be a point where ADS-equipped vehicle deployment changes overall motor vehicle safety outcomes and safety performance. Flexibility to changes in safety metrics may need to be considered as ADS-equipped vehicle deployment and technology changes.
- Event data recording related to ADS-equipped vehicles and crashes may need to be tracked in a manner that complies with applicable data privacy laws. Data collection could consider technology compatibility and changes. Appropriate incident data should be transparent to the public and other stakeholders.
- ADS-equipped vehicle safety considerations could examine learnings from the aviation industry, which has given thought to various human factors needs, such as providing a button for cabin occupants to request help.

3.2 Is there a need for widely-adoptable measurement methods for ADS-equipped vehicle safety? Are there risks in not pursuing such methods? If so, what are some examples?

The attendees in the four breakout groups were in agreement that there is a strong need for widely-adoptable methods to measure the safety of ADS-equipped vehicles with the ultimate goal to ensure safe deployment of ADS-equipped vehicles and increased public trust and adoption.

In developing the new safety measurements, many attendees believed that the methodology used to do so becomes a primary concern. These attendees asserted that it would be beneficial if existing vehicle measurements and test can be built upon and used as a baseline rather than starting anew. It was pointed out that this approach is consistent with established industry practice, and analogous to the approach taken in crash safety standards. Some breakout attendees advocated further development of NHTSA's Voluntary Safety Self-Assessment (VSSA). However, it was noted that an independent assessment, with transparency, may provide the public with additional confidence in the safety of ADS-equipped vehicles, thus facilitating faster adoption.

Attendees identified that some of the factors in the development of measurement and safety methods are rigor and robustness with little room for gaming the test. Trust was discussed as critical for the

[17] The RAND report, referenced in footnote 4, addressed the use of near crashes (along with instantaneous safety metrics) in listing existing measurement ideas relevant to the proposed, integrated concept it called "roadmanship" in Table 3.2 on p.33. https://www.rand.org/pubs/research_reports/RR2662.html

acceptance of ADS technology and having a widely accepted method for measuring safety may help support building this trust. It is entirely acceptable for a vehicle to be designed to pass a test, as long as the test is representative of real-world deployment, thus providing a means for assurance that the system is safe with respect to that use case. Since establishing public trust emerged as a common goal, the safety test results were recommended to be made public.

Considering that the measurement methods may include ADS-equipped vehicles with features at SAE Level 3 and higher, where each vehicle has a separate ADS for each Operational Design Domain (ODD), some participants stated that the measurements and tests could be tailored to specific ODDs. It was pointed out that the ODDs vary with respect to roadway types, geographical location, speed range, lighting conditions for operation (e.g. day or night), weather conditions, and other operational domain constraints. Geofencing was proposed as one approach to ensuring that ADS-equipped vehicles do not stray beyond the ODD. It was also noted that different safety measures may be required for vehicles with different applications such as goods delivery or passenger movement. Some attendees believed that passenger vehicles may be subject to more tests than goods vehicles to ensure the safety of the occupants of vehicles (e.g., safe termination of ride, emergency stops). Other participants noted that goods delivery ADS-equipped vehicles will nevertheless be interacting with pedestrians and occupants of other vehicles. Additionally, many agreed that there should be more discussion around how to keep the delivery vehicle safe during an incident. Some attendees stated that, for legacy vehicles, while designers continue to fall back on holding the human supervisor accountable for Level 3 vehicles, with many ADS-equipped delivery vehicles, a human is not inside the vehicle to steer it to safety.

Some participants noted that an integral part of any test lies in its governance — which body(ies) are in charge of developing, adopting, and enforcing the test. Today, state governments play an important role in facilitating the deployment and on-road testing of ADS-equipped vehicles as they are responsible for vehicle licensing and registration, traffic laws and enforcement, and motor vehicle insurance and liability. However, a shared objective that emerged from this workshop is the need for the establishment of a consistent national framework rather than a patchwork of incompatible state-level tests to allow vehicles to easily cross state lines. Some believed that this could be accomplished by building a federal menu of specific safety tests that cities or states can select based on applicability and desirability. Others believed that a framework could be desirable, where the federal agency sets the national standards while states can elect out of specific provisions.

To conclude, while the breakout attendees agreed on the need for widely-adoptable measurements for safe deployment of ADS-equipped vehicles, the concerns, challenges, and potential solutions outlined above should be reviewed to enable robust measurements and assessments.

3.3 What are possible safety measurement methods (simulation, test track, on-road, etc.)? What are possible safety metrics (miles driven, pass/fail vs. formal model, etc.)?

Workshop attendees, including presenters, introduced and discussed possible safety metrics and methods including specific suggestions and general attributes. Several attendees suggested reordering of these two questions to address metrics first and then methods. Attendees agreed that simulation, test track, and on-road testing may all be necessary for the development of self-driving technology. Some suggested that a minimum set of requirements and guidelines could be established in order to compare and contrast various OEM’s test results. Furthermore, some attendees would like to see federal and state regulators use the initial metrics and that they could help create and review the evolving metrics based on changing ODD or commercial landscape. It was noted that some experts propose a metrics approach that includes tying both leading and lagging metrics directly to the claims, argumentation, and evidence in a system-level safety case.

Regarding the approach to metrics, attendees made the following observations:

- Metrics should be ODD based;
- Some metrics are binary while others have a larger range of possible values;
- Some metrics are leading and others lagging indicators;
- Metrics could be used by federal and state regulators;
- Metrics should have associated criteria (i.e. should be a validated set of metrics);
- Metrics should evolve based on changing ODD or commercial/regulators landscape; and
- Metrics should not hinder technology advancements.

The general consensus amongst breakout attendees was that metrics should drive methods and they should be framed by ODDs. Metrics should not be overly restrictive and should encourage technology advancement. Attendees also suggested that the metrics should support initial and subsequent deployments. Some metrics may reflect a pass or fail condition where some indicate a degree of accuracy that can be evaluated based on other criteria. Metrics should be aggregated to support the technology throughout the entire life-cycle.

Participant discussion included a number of possible metrics:

- Harm to external things;
- Stress level of occupants;
- Miles driven by an ADS in a representative set of ODDs (w/safety driver);
- Proximity or safe distance;
- Performance compared to a human;
- Time to Collision (TTC); and
- Hard Braking.

Attendees generally suggested that the performance and functionalities of an ADS-equipped vehicle should be better than a human-driven one on the road today. It was noted by attendees that other road users are accustomed to a certain behavior for vehicles in terms of their interaction with other vehicles, pedestrians, public safety agencies (e.g., law enforcement, fire, and EMS), and their surrounding areas in general. ADS-equipped vehicle developers should be aware that new behaviors, unknown or counter to normal and law-abiding humans or human-driven vehicles may jeopardize public safety or hinder the technology adoption. Some attendees also acknowledged the difficulty in having the appropriate data available to do this comparison.

Participants also discussed views around some of the methods that could be used to generate metrics. The methods ranged from physical and virtual testing to how these methods would need to be sensitive to how ADS are developed.

With respect to methods, attendees provided input that:

- Methods should be driven by required metrics;
- Simulation, test track, and on-road tests are all appropriate and necessary (some participants provided input that simulation is most appropriate for development, while there is concern about its use due to the lack of uniformities in what is simulated and how);
- Minimum set of requirements and guidelines could be established for simulation – there needs to be the ability to uniformly compare test results;

- Methods need to take into account software upgrades (including additional training for artificial intelligence (AI) technology used in the ADS); and
- Methods need to take into account ODD changes and extensions.

Attendees also mentioned best practices related to ADS safety measurement in areas such as vehicle design attributes, software design attributes, and process expectations. Other issues that were raised included:

- ADS should be demonstrated as safe before operation;
- Standards could be defined for safety self-assessments that could be made available to the public;
- Responsibility Sensitive Safety (RSS) is a possible method to evaluate further;
- Redundancy may be an important consideration;
- ADS should be able to extract itself from hazardous situations; and
- Appropriate software design processes should be used, including adaptive processes that learn from deployments.

3.4 Are there emerging best-practices around pre-deployment safety measurement methods? Around post-deployment measurement methods? (including the methods and metrics described above).

The breakout attendees agreed that there are some nascent best practices emerging for ADS-equipped vehicle pre-deployment and post-deployment safety measurement but that a significant effort is needed to flesh these practices out, to augment them, and to combine the results into a comprehensive safety measurement methodology. It was further noted that there is a coupling between pre- and post-deployment measures that will need to be considered.

For pre-deployment, the USDOT has recommended the VSSA process in which developers are encouraged to disclose detailed descriptions of their safety measurement practices. As noted during a workshop presentation from USDOT, a significant cross-section of developers has posted VSSAs on their web sites, however, these documents were not reviewed during the workshop.

There has, to date, been very limited prototype testing of SAE Level 4 and Level 5 ADS-equipped vehicles (L4 and L5 were the focus of the workshop). As a result, most of the discussion in the breakout sessions of the workshop involved brainstorming about the safety measurement practices that could be used post-deployment. It was noted that it is important to share candidate methodology(s) with the public and media for feedback throughout the safety measurement development process to help establish trust and confidence.

3.4.1 PRE-DEPLOYMENT

Breakout attendees suggested a combination of off-road and on-road testing as a pre-deployment safety measurement methodology. Simulation was mentioned as a possible third component, though there was concern expressed that it is difficult to validate simulation environments. It was acknowledged that on-road testing with safety-assessment drivers was a useful, existing practice. Beyond this, the following practices were suggested:

- Developing a safety test-case framework:
  - Use of such a framework would require developing dynamic test facilities, or test tracks, capable of exposing ADS-equipped vehicles to the breadth of conditions they may need to handle.
  - Attendees noted that a number of test case frameworks and driving scenario databases are being developed and noted:
    - Efforts by Waymo and Aptiv, noting that Uber and Lyft are likely to also contribute scenarios; and
    - Low-probability high consequence events are not likely to be captured in such a database.
- Attendees noted that for deploying standardized safety testing that:
  - Standardized safety testing is likely to be mandated by regulatory agencies to validate developer testing and self-certification; and
  - Even in standardized safety testing, there may be variability in testing requirements across the country due to variations in operational design domain (ODD), actors, municipalities, etc.
- Attendees mentioned these examples of possible pre-deployment safety metrics:
  - Number of miles driven, and number of disengagements, in a representative set of ODDs.
  - Resilience – we may need to test the ability of ADS-equipped vehicle systems to recover and the time required for recovery when faced with anomalies.
- Several examples of needed research efforts on the topic of pre-deployment safety measurement were noted:
  - Definitions of safe driving behavior need to be established. Manufacturers may need to understand what it means to test their vehicles
    - across representative infrastructures;
    - against human drivers;
    - against their competitors' ADS-equipped vehicles.
  - Human driving excellence needs to be defined if developers and society are to use this as a goal for ADS driving. This could help in establishing ADS-equipped vehicle safety requirements/benchmarks. Responsibility Sensitive Safety (RSS) was mentioned as one possible beginning point for such a definition.
  - Comparison of human perception with the capabilities of sensors – one approach to such a comparison was by breaking down the driving task into its different components and comparing the performance of sensors with human perception for each task.
- Additional remarks by workshop attendees included:
  - Human drivers in the future may need to test their driving abilities against those of ADS-equipped vehicles.
  - Best practices from existing automobile safety methodology and standards and from other transportation domains where applicable can be considered.
    - Each of the elements of the ISO standards should be examined and applied to address ADS-equipped vehicle pre-deployment safety measurement needs.
    - The aviation sector has MOPS - minimal operational performance standards – and the equivalent could be developed for ADS safety.

3.4.2 POST-DEPLOYMENT

For post-deployment, workshop attendee focus was on approaches using metrics for assessing ADS-equipped vehicle safety performance:

- Concern was expressed about the implementation of post-deployment updates such as software updates and the need for a post-deployment safety measurement methodology that would guarantee their safety after update;
- Further, it was suggested that real-time continuous monitoring/testing would need to be deployed to assess the health of vehicles and to trigger safety maintenance and re-assessments as required.

Additional attendee comments on examples of post-deployment safety metrics included the following.

- Crashes, traffic incidents, and traffic violations including near crashes.
  - These metrics will need to be ODD-Based.
- Comparative measurement against human drivers' performance.
- Use of human best driving practices to establish safe driving metrics.
- Measurements derived from consumer complaint databases and private databases, although maintaining and processing such data would be a major undertaking and would require the development of appropriate tools. Database examples that were cited included:
  - NHTSA consumer complaint database with 2 million complaints;
  - "How's my Driving?" consumer incident databases; and
  - Crowd sourced inputs – in particular validated or curated databases.
- System-wide behavior metrics are needed, (e.g., assessment of ADS-equipped vehicle interactions with infrastructure, each other, human drivers, pedestrians).

3.5 Can measurement of human response to ADS-equipped vehicle safety be a part of the calculation and, if so, in what way?

The response the workshop participants gave to this question was affirmative but conditional based on scope. To frame this question, workshop attendees chose to focus on Levels 4 and Level 5 ADS features where the human driver was not involved but the ADS-equipped vehicle users are considered occupants and not drivers. With that scope defined, the next delineation in thought was between human interaction inside the ADS-equipped vehicle as well as vehicle interaction with human occupant(s) and other road users.

Concerning the occupant(s) inside the vehicle, discussions centered around allowing the occupant(s) the ability to initiate executive decisions to manually override the vehicle in necessary situations (i.e., start minimal risk condition procedures). However, questions that are still open and to be grappled with include:

- Should occupants be allowed to execute certain decisions?
- Which situations are considered critical?
- Can the ADS-equipped vehicle override the human decision?

Participants noted that work is taking place within SAE International's On-Road Automated Driving (ORAD) Committee for development of safe pull over[18] and emergency stop standards.[19]

Some attendees asserted that the more pressing and potentially larger concern about ADS-equipped vehicles and human interaction was the seemingly limitless interactions possible outside the vehicle. Two general perspectives emerged from the breakout attendees on how to address this problem. The first perspective considers Levels 4 and Level 5 ADS-equipped vehicles as any other new transportation or infrastructure feature in which humans will figure out de facto rules of interaction with the new technology. The second perspective is that ADS-equipped vehicles should have the ability to predict human behaviors and interactions with the vehicle and appropriately respond. Without taking a stance on either of these positions, it would be appropriate to say that whichever direction is chosen going forward, attendees agreed that more research on the behaviors and interactions of humans and ADS-equipped vehicles needs to be undertaken.

The misbehavior of humans, intentionally or otherwise, was mentioned several times by participants. For example, human occupants may inappropriately activate vehicle controls. Also, humans, inside and outside of the vehicle, may deliberately game or "test" the ADS-equipped vehicles' stop mechanism. This is currently an active research area. An example provided by a participant is a study on interactions of pedestrians and vehicles in Miami.[20]

Additional remarks included the following:

- A critical function of human response to ADS-equipped vehicles is the interaction with occupants and road users with disabilities (i.e., occupants that are limited-vision or blind, hearing impaired or have wheelchair accessibility requests);
- Another critical function is law enforcement's ability to regulate ADS-equipped vehicle behavior through stopping and ticketing procedures, especially when there are no human occupants in the vehicle; and
- The ADS brand and reputation may influence the perception of road users towards an ADS-equipped vehicle when sharing the road. This discussion mirrored a concurrent social media discussion on trustworthiness of ADS-equipped vehicle providers.[21]

[18] "Guidelines for Safe On-Road Testing of SAE Level 3, 4, and 5 Prototype Automated Driving Systems (ADS)" SAE On-Road Automated Driving Committee (ORAD), J3018_201503, Issued - 13 March 2015. https://www.sae.org/standards/content/j3018_201503/

[19] "Taxonomy and Definitions for Terms Related to Automated Driving System Behaviors and Maneuvers for On-Road Motor Vehicles" SAE On-Road Automated Driving Committee (ORAD), J3164, Work in Progress - 31 January 2018. https://www.sae.org/works/documentHome.do?docID=J3164&inputPage=wIpSdOcDeTallS&comtID=TEVAVS

[20] "When Self Driving Cars Meet Florida Drivers" Small, Andrew. CityLab, 4 December 2018, https://www.citylab.com/transportation/2018/12/ford-argo-self-driving-cars-autonomous-vehicles-testing-miami/577303/

[21] "What do you need to know about an ADS-equipped vehicle provider to accept it as trustworthy?" @riellybrennan (Reilly Brennan) et al. Twitter, 23 June 2019, https://mobile.twitter.com/reillybrennan/status/1142930265737584640

Participants expressed that there should be clear and predictable communication and signaling of intent between the ADS-equipped vehicles and their occupants, as well as with other road users and pedestrians outside the ADS-equipped vehicle.

3.6 What are possible next steps?

There was a clear consensus amongst the workshop attendees that the ADS-equipped vehicle community would benefit greatly from a community-wide effort to establish a coherent, widely-supported, comprehensive safety methodology framework that would help focus the efforts of the various ADS-equipped vehicle stakeholders to advance the introduction of ADS-equipped vehicle technology. However, there was still considerable disagreement about the details of any framework.

However, attendees emphasized that, to be successful, the effort must engage stakeholders across the ADS-equipped vehicle community. Participants identified stakeholders to include: designers; developers; manufacturers, component suppliers; standards and testing/certification organizations; transportation services providers; the insurance sector; first responders, law enforcement, and public safety officials; local, state and federal agencies; academic and industrial researchers; safety and professional organizations; consumers; and the general public.

Workshop participants described what they believe are key elements that should be included in any further community-wide effort to develop a framework as follows.

First, attendees asserted that any community wide effort should use current practices and ongoing activities as the starting point. Attendees believed that members of the community should work together to define a process, including roles and responsibilities, that the community can embark upon and focus on how far the community can achieve progress over the next 6 months to a year. Participants described efforts that could be undertaken by community members: work together to produce, as needed, documents that progressively capture the desired empirically-based safety methodology framework; and iteratively evolve, extend, and perfect the framework over time. The effort should also capture and evaluate what has already been developed across the community and roll it into frameworks, as well as define the minimal requirements the community agrees on and use that as a basis for advancement. Finally, participants suggested that the effort should engage the public early and often to foster confidence in ADS-equipped vehicles.

Second, attendees asserted that the effort should describe a basic strategy for safety measurement and metrics. It was believed that this strategy could (1) define and consider the impact of ODDs; (2) identify both core and edge-case scenarios; and (3) develop objective metrics that allow safety to be effectively measured for core and edge scenarios. This strategy could also include a methodology for measuring and processing near crashes and consider what the role of infrastructure could be in establishing an optimized safety methodology.

Third, attendees added that a community-based effort could be used to assist in defining the data needed to support the framework. For example, the effort could examine what data is relevant for measurement and performance (e.g. what is meaningful? How could it be collected for both pre- and post-deployment data and different event types? The effort could also analyze data ownership and access provisions and assess needs for a repository for the community for relevant data.

Fourth, attendees indicated that the effort could assist in outlining potential testing and certification strategies. That is, the effort could be useful to government authorities in assisting in analyzing existing testing and certification strategies for applicability gaps, and opportunities. The effort could also assist in evaluating approaches for public acceptance and trust.

There was broad consensus among workshop participants that the time is right for a compact, focused, community-wide effort of the type described above. Participants were generally optimistic that such an effort could be successful in creating an initial ADS safety measurement framework that would be useful to the community. Participants also noted that an initial framework and the process used to create it could be a starting point for a continuing process for evolving and improving safety measurement approaches across the sector as experience is gained, technology advances, and consumer acceptance expands.

4 Appendix A - Abbreviations

AAMVA – American Association of Motor Vehicle Administrators
AASHTO – American Association of State Highway and Transportation Officials
ADAS – Advanced Driver Assistance Systems
ADS – Automated Driving Systems
AI – Artificial Intelligence
AMP – Automated Mobility Partnership
CAT – Cooperative Automated Transportation (Coalition)
CPS – Cyber-Physical Systems
DDT – Dynamic Driving Task
DNN – Deep Neuro Networks
EMS – Emergency Medical Services
FMVSS – Federal Motor Vehicle Safety Standards
FO – Fail Operational
FS – Fail Safe
IIHS – Insurance Institute for Highway Safety
ISM – Instantaneous Safety Metric
ISO – International Standards Organization
MOPS – Minimum Operational Performance Standards
NDD - Nationalistic Driving Data
NHTSA – National Highway Traffic Safety Administration
NIST – National Institute of Science and Technology
ODD – Operational Dynamic Domain
OEDR – Object and Event Detection and Response
OEM – Original Equipment Manufacturer
ORAD – On-Road Automated Driving (SAE Committee)
RSS – Responsibility-Sensitivity Safety
SAE – Society of Automotive Engineers
TTC – Time-to-Collision
UNECE – United Nations Economic Commission for Europe
USDOT – U.S. Department of Transportation
VSSA – Voluntary Safety Self-Assessment
VTTI – Virginia Tech Transportation Institute
WP – Working Party

5 Acknowledgement

We acknowledge the contributions to this report by the workshop participants, who represented the broad spectrum of stakeholders in automated driving safety. In particular we acknowledge the volunteers that assisted in writing and reviewing the drafts.

The writing team volunteers were responsible for taking notes during the workshop and developing the various initial drafts for this workshop report. The writing team included Sam Afsoos, Nat Beuse, Myra Blanco, Michelle Chaka, Christopher Greer, Edward Griffor, Greg Leeming, Kenneth Leonard, Tom Linn, John Maddox, Nick Royal, Omar Sheikh, Prachi Vakharia, Jack Weast, Simone Hadley Wilson and David Wollman.

Several others volunteered to serve on the 'red team' that reviewed and provided additional inputs to the draft produced by the writing team. The red team included Jackie Erickson, Khalid Halba, Aaron Jones, Shawn Kimmel, Sean Laffey, James Li, Matthew Magsamen, Ann Ngo, Aviral Shrivastava and Edward Thai.