EDGE CASE RESEARCH

Comments on Framework for Automated Driving System Safety

NHTSA-2020-0106-0001

March 31, 2021

Comments of Philip Koopman, Ph.D.
CTO & Co-Founder, Edge Case Research
Associate Professor Carnegie Mellon University

I welcome the opportunity to comment on this ANPRM document. It is indeed good to see US DOT and NHTSA exploring potentially ADS-specific governmental safety frameworks that go beyond the FMVSS approach.

My perspective in writing these comments is informed by my experiences. I have been involved in ADS safety for approximately 25 years, starting with a role on the Carnegie Mellon University Navlab team for the Automated Highway System (AHS) project in the 1990s. I have worked with industry teams using a variety of safety standards, performed design reviews, and otherwise been involved in domains including conventional automotive, ADS equipped vehicles, rail, chemical processing, industrial controls, electrical power, building controls, vertical transportation, consumer products, medical systems, aviation, and submarine combat systems. I was the principal author of the content in ANSI/UL 4600, and also serve on other relevant SAE-affiliated ISO safety standard committees. I currently teach an annual course on embedded system software quality, safety, and security at Carnegie Mellon University, with significant coverage of ADS related topics in both my teaching and research. However, these comments express solely my own opinions and not those of the University.

Summary of High Level Recommendations:

1. **Industry standards.** NHTSA should encourage conformance to normative safety standards written by the automotive industry and stakeholders themselves and issued by accredited Standards Development Organizations, including but not limited to ISO 26262, ISO 21448, ANSI/UL 4600, and safety-relevant security standards.

2. **Transparency.** NHTSA should act to increase transparency with regard to safety in the automated vehicle industry.

3. **Safety First.** NHTSA should encourage the industry to collaborate on safety and compete on factors other than safety.

4. **Human Operators.** NHTSA should ensure that the division of tasks between human operators and automated vehicles results in acceptable safety, and does not place unreasonable demands on human capabilities.

5. **Safety Cases.** Longer term, NHTSA should transition from a test-based posture to a safety case-based posture that includes testing as a component.

6. **Safety Critical Computer System Skills.** NHTSA should significantly increase their staffing strength in computer-based system skills, especially in the area of software.

Responses to questions:

Question 1:

Describe your conception of a Federal safety framework for ADS that encompasses the process and engineering measures described in this document and explain your rationale for its design.

A Federal Safety Framework for ADS should encompass the following elements. These elements have been selected to make use of existing industry efforts and provide a level playing field for an implementation-neutral approach to establishing a baseline for and continually improving safety.

1. **Industry standards.** NHTSA should encourage conformance to safety standards written by the automotive industry and stakeholders themselves, and issued as normative standards by accredited standards organizations (e.g., ISO, ANSI/UL, SAE). This includes but is not limited to ISO 26262, ISO 21448, ANSI/UL 4600, and safety-relevant security standards.

    a. This should include full self-disclosure of standard conformance status for every highly automated vehicle operating on public roads, including aspects of the vehicle for which conformance is declared. (The sole exception should be test vehicles under the immediate control of a qualified safety driver as part of a publicly declared testing effort.) This would not necessarily be a requirement for conformance, but rather a requirement to be transparent and forthcoming about conformance with industry-created standards (or lack thereof). If no safety standards are conformed to, that should be so stated. A clear and unambiguous statement should be required (e.g., “we conform to ISO 26262”) rather than a vague statement such as “we use approaches inspired by [standard]” or “we adopt techniques drawn from [list of standards].”

    b. It is important to note that such self-disclosure does not require public disclosure of sensitive proprietary technical information. For example, conformance to ANSI/UL 4600 does not require disclosing any technical information to any organization external to the organization declaring conformance.

    c. I note that in industries other than automotive there is either required or voluntary conformance with comparable domain-specific safety standards. It is difficult to understand how the ADS industry, which justifies its need for regulatory breathing room by promising to make things safer, can at the same time fail to follow industry consensus safety standards for applicable aspects of their vehicles.

2. **Transparency.** NHTSA should act to increase transparency with regard to safety in the automated vehicle industry.

a. Specific steps should include updating the NHTSA-defined VSSA guidance scope to include all major aspects of ANSI/UL 4600 compared to the current subset of topics covered. (In fairness, the VSSA guidance was created before the April 2020 issuance of ANSI/UL 4600, so this should be considered an evolution of the VSSA guidance to track evolving issued industry standards.)

b. NHTSA should also increase industry participation rates in releasing technically substantive VSSAs. A properly formed VSSA document should in fact be a high level but technically substantive disclosure of the relevant safety case, and should be issued by every company putting a vehicle on public roads. This should include companies testing on public roads publishing a VSSA scoped to address the safety of the testing effort.

c. The release of some recent, technically substantive VSSAs and the public Web posting of the Uber ATG safety case framework demonstrate that significantly more transparency is viable without undue disclosure of sensitive proprietary information.

d. NHTSA should define and strongly encourage reporting safety outcomes (lagging metrics) in a uniform and transparent manner to demonstrate via data that ADS technology results in safer roads. This information should be supplied by manufacturers and operators rather than solely relying upon, for example, police reports. (Note that the industry itself could drive this standardization; it need not be a NHTSA-defined standard.)

e. A specific concern is ensuring that potential safety issues in one mode of operation (e.g., driver supervision) should not be buried in aggregate data (e.g., by mixing less safe mode data with safety improvements from active safety features during manual driving). An additional concern is that metrics should drive improved safety for road users rather than be used as a score card that is gamed to show progress in a “race to autonomy” (e.g., disengagement metrics are problematic for this reason).

3. **Safety First.** NHTSA should encourage the industry to collaborate on safety and compete on factors other than safety.

a. Safety should be a given. As with the airline industry, achieving industry-wide safety should involve cooperation among all stakeholders. NHTSA is in a unique position to foster such cooperation, potentially with support from neutral organizations.

b. A starting point can be a shared repository of potential hazards to be addressed when relevant to an ADS-equipped vehicle’s ODD.

c. NHTSA should facilitate a dialog on the topic of how safe is safe enough, including all stakeholders. This should address issues such as relevant metrics,

risk transfer, taking credit for safety improvements to offset higher-risk operating modes within vehicle fleets, and degree to which near term risk can (or even should) be traded off against potential long-term aspirational safety improvements.

d. A longer term goal should be a set of ODD-specific lagging metric safety performance indicators and baseline minimum targets based on human driver performance to set a level playing field for safety performance reporting and outcome assessment.

4. **Human Operators.** NHTSA should ensure that the division of tasks between human operators and automated vehicles results in acceptable safety.

a. This should include monitoring deployed vehicles for an unsafe division of responsibility (e.g., systems overly prone to automation complacency that results in elevated mishap rates) as well as longer term research into driver monitoring effectiveness at ensuring operational safety.

b. NHTSA should encourage the industry to develop standards for measuring driver engagement in the context of driver monitoring systems and their effectiveness in naturalistic driving situations.

c. NHTSA should address all outstanding NTSB recommendations, especially in the area of driver engagement. (See: https://www.regulations.gov/comment/NHTSA-2020-0106-0617)

5. **Safety Cases.** Longer term, NHTSA should transition from a test-based posture to a safety case-based posture that includes testing as a component.

a. For some aspects of safety, a test-centric approach is appropriate. However, in essentially all areas of large-scale computer-based system safety, testing is necessary but insufficient to ensure acceptable safety. Given the unique nature of machine-learning based technology incorporated into typical ADS equipment, process-based metrics and leading indicator metrics based on field engineering feedback will be essential to demonstrate and improve safety over the course of deployment.

b. A safety case-based NHTSA posture should involve asking ADS-equipped vehicle makers to use safety cases and (a) define what they mean by safe, (b) explain what reasoning is being used to argue they are safe, and (c) explain the basis of evidence to support that reasoning.

c. A critical part of this will be to ensure not only that ADS equipped vehicles send back field data to ensure that the safety case is valid in practice, but also that a metric-based approach ensures that the ADS design and deployment

organizations are actually paying attention to and taking action upon data that indicates potential safety issues before loss events occur.

d. While good engineering, sound data collection practices, simulation, closed course testing, and safe road testing will all play a part in ensuring safety, the precise role of each of these is still open for ADS technology. Therefore, NHTSA should concentrate on ensuring that manufacturers have a coherent story to tell about safety rather than mandating what that story actually is. ADS equipped vehicles should only be deployed when they are demonstrably safe, but the form of the demonstration (which will need to include more than driving an actual vehicle) should be informed by the specific safety case involved.

6. Safety Critical Computer System Skills. NHTSA should significantly increase their staffing strength in computer-based system skills, especially in the area of software.

a. NHTSA has historically under-staffed in the area of computer-based system safety, and especially software safety. However, in recent years automobiles have transformed from electromechanical systems to computers-on-wheels. Especially in electric vehicles, there is simply no way to understand whether a vehicle is acceptably safe without understanding computer technology.

b. Currently, NHTSA reports routinely do not rule in computer-based system defects (and especially software) when considering potential root causes of mishaps. Yet there is a dramatic rise in software-related recalls. The writing is on the wall: significantly more capability is required in the area of safety critical software if NHTSA wants to remain relevant to actual safety outcomes. It is recognized that budgets are limited, but this is an area that simply cannot be neglected.

Question 2:

In consideration of optimum use of NHTSA's resources, on which aspects of a manufacturer's comprehensive demonstration of the safety of its ADS should the Agency place a priority and focus its monitoring and safety oversight efforts and why?

FMVSS-style tests should continue to serve to ensure some minimum level of vehicle performance competence and physical vehicle capabilities. Any FMVSS amendments or waivers should preserve the safety purpose of the test rather than emphasizing the form of the test. For example, low tire pressure tests should emphasize ensuring a viable mechanism is in place to detect and report under-inflated tire situations to a person responsible for correcting the situation, rather than requiring a particular warning light approach that might be unlikely to provoke a suitable corrective response, especially in an uncrewed vehicle.

NHTSA should not spend massive resources attempting to define a comprehensive ADS “driver test.” While a minimalistic road competence test could potentially keep unsophisticated design teams that are not capable of fielding safe vehicles off the roads, a very extensive “driver test” would consume huge NHTSA resources and would be unlikely to provide strong evidence of operational safety at even the level of average human driver ability. At best such tests would only be likely to identify ADS designs so bad that they can’t pass a predefined test. (That goal is not a bad one, but could and should be achieved in an economical manner.) Such a test might, however, provide protective cover for an organization that is motivated to cut corners on safety by building to the test (even a randomized test can be expected to be gamed if it is the only safety measure required) instead of actually building a vehicle that will be safe in the real world.

Given limited resources, instead of prioritizing road tests, NHTSA should prioritize ensuring transparency and a level playing field for achieving acceptable operational safety as outlined in the response to Question 1.

Question 3:

How would your conception of such a framework ensure that manufacturers assess and assure each core element of safety effectively?

Due to the dramatic change in responsibilities for and even the removal of human drivers, the introduction of ADS technology serves as an inflection point in automotive safety practices. A healthy safety culture for the ADS industry can and must be established without waiting for loss of life as motivation. NHTSA has the opportunity to play a pivotal role in this process.

NHTSA will make much more effective use of its limited resources if it can serve as a quality check on a healthy industry set of safety practices rather than being relegated to a sole role of the "safety police." To that end, NHTSA should do everything it can to foster a healthy ADS industry safety culture via voluntary means, and wield the lightest weight practical pressure to accomplish this. The highest leverage is available in ensuring safety culture is working rather than designing and running tests that the manufacturers should already be defining and executing on their own. (To be sure, I am not advocating for eliminating NHTSA's crucial monitoring and enforcement role. Rather, I am pointing out that proactive safety culture improvement is a cost-effective way to avoid needing to exercise the enforcement role, and is likely to yield an overall better result than a purely test and recall role.)

It is important not to confuse how an ADS system operates with how it is made safe. For example, the required ability of a specific sensor to detect a specific type of target at a specific distance will vary depending upon the ODD as well as the role of that sensor in the specific system design. (Example: effective road obstacle detection range required for a low speed shuttle will differ dramatically compared to a highway speed heavy truck.) Acceptable safety might be achieved with different sensor suites and sensor capabilities, and with significantly different internal processing architectures, especially when considering a diverse range of ODDs.

Rather than attempting to assess functionality beyond basic tests in the vein of current FMVSS and NCAP approaches, NHTSA should instead emphasize assessing whether an organization has created a viable safety case, has performed self-determined tests responsive to that safety case adequately, and whether organizations are indeed paying attention to and taking action upon emergent safety issues of their own accord.

As an example, when a safety relevant vehicle defect is detected via NHTSA field data surveillance, the manufacturer should not only issue a recall to fix the problem, but should also explain to NHTSA: (a) why they missed the early warning signs of this problem and therefore did not fix the issue before it rose to the level of a recall discussion, and (b) how they are changing their safety case to avoid similar problems with both product and process in the future. Such a process should be non-punitive if the manufacturer is acting in good faith, and should require that manufacturers proactively gather their own data and perform their own corrective actions without waiting for NHTSA to act. In other words, NHTSA should serve as a check and balance on the industry, not as the primary finder of defects.

If core element evaluation is desired, it should be driven by confirmation of properties required to satisfy the manufacturer's safety case. This topic is covered by ANSI/UL 4600 section 16 Safety Performance Indicators. In brief, such indicators are metrics designed to detect violations of claims made in the safety case (i.e., a metric used to ensure that claims are actually valid in operation, and to flag instances in which safety case claims somehow become false in practice due to changing conditions or unforeseen gaps in the safety case arguments). Until the time comes that the industry settles on a small number of standardized safety case templates, leading safety metrics will need to be responsive to the differences among safety cases.

Question 4:

How would your framework assist NHTSA in engaging with ADS development in a manner that helps address safety, but without unnecessarily hampering innovation?

This question is addressed in terms of the six overarching themes of these comments:

1. **Industry Standards.** Conformance to standards created by the industry itself should not hamper innovation. In particular, ANSI/UL 4600 is specifically designed to be technology neutral and to permit designers flexibility in only taking action upon considerations that are relevant to their technology and ODD. ISO 26262 and ISO 21448 define tailorable processes and methods rather than technology-specific implementations, and do not put any substantive constraints on safe innovations.

2. **Transparency.** Asking ADS designers to explain why they are safe enough for public deployment is a question they themselves must be able answer before they can responsibly deploy. Explaining why they are safe need not constrain innovation nor expose sensitive technology trade secrets. Any explanation that amounts to "trust us, we're smart and we work really hard" is a red flag indicating a serious lack of engagement with regard to transparency in public safety. Claims of trade secret protection should not be used to dodge accountability for public safety.

3. **Safety First.** An initial emphasis on transparency and a NHTSA role in establishing a combination of lagging metrics and transparency on feedback loops to correct problems permits safe innovation. Only unsafe innovation would trigger a significant violation of lagging metrics.

4. **Human Operators.** Establishing and reporting lagging metrics for safety related to human operator interaction with ADS equipped systems permits safe innovation. Research into good and bad practices will promote innovative good practices.

5. **Safety Cases.** Safety cases permit responsible innovation by decoupling the "how" from the "what" in terms of what safety goals are established and achieved.

6. **Safety Skills.** It is unreasonable to expect NHTSA to be effective at ensuring safety of software-intensive life critical systems without having robust staffing covering software safety skills. Continued weakness in this area simply sets up NHTSA for failure.

Additional specific recommendations are:

NHTSA should reformulate the VSSA guidance to encompass the high level topics in ANSI/UL 4600 and any additional topics identified as important to ADS-equipped vehicle safety (e.g., mitigation of fallback driver complacency in SAE Level 3 vehicles). This would provide a more complete picture of safety for those organizations who submit a VSSA, while rooting the VSSA contents in an industry-created ANSI standard developed primarily for ADS equipped vehicles. This would build upon the existing VSSA process to further improve NHTSA engagement with manufacturers on the topic of safety.

NHTSA should create a rubric for scoring the completeness of VSSA submissions in terms of degree to which they address all essential aspects of the VSSA guidance at a level of detail more fine-grain that the approach of providing a specific set of chapter topics. This rubric should not be intended to “grade” safety itself, but rather the completeness of the disclosure made by the VSSA submission. For example, do VSSA submissions unambiguously state their standards conformance posture? Such a rubric could encourage more complete and comprehensive disclosures with technical substance rather than high level marketing brochure style documents.

NHTSA should encourage ANSI/UL 4600 conformance for companies building and deploying ADS equipped vehicles. It is important to note that self-certification with no disclosure of sensitive technical information is explicitly supported by the ANSI/UL 4600 conformance model. Encouraging ISO 26262 conformance as well as ISO 21448 conformance are also desirable, but in the context of an ADS discussion, ANSI/UL 4600 is uniquely positioned by virtue of it being an ADS-specific standard. As a practical matter, conformance to ANSI/UL 4600 will often involve some level of conformance with ISO 26262 and ISO 21448 as well for vehicles that operate on public roads.

Question 5

How could the Agency best assess whether each manufacturer had adequately demonstrated the extent of its ADS' ability to meet each prioritized element of safety?

If pre-release testing is desired, a manufacturer should already have test procedures in place to ensure that they are acceptably safe in accordance with their safety case. Replication of selected or randomly chosen such acceptance tests via witness testing can provide some degree of confidence in safety. The witness could be from NHTSA or a qualified and accredited third party bound by a technology NDA but permitted to accurately report pass/fail for witness testing purposes. (A robust ecosystem of such third party entities already exists for other purposes, such as certifying conformance to ISO 26262 and other safety standards for other industries.) This would avoid NHTSA spending significant resources designing their own tests.

In other words, if manufacturers claim they are safe, they must have some basis for that claim. NHTSA could simply ask to check whether the data supporting the manufacturer claims can be replicated on their own terms. Such an approach should be coupled by field engineering feedback so that a company that performs overly simplistic tests will be confronted with real-world data of any inadequacies of their product before significant losses in real world operations are likely to have occurred. While there are obvious limitations to such an approach, it is clearly better than allowing manufacturers to use completely opaque processes to design and deploy systems without any checks and balances on computer-based system and software safety at the time of deployment.

Question 6:

Do you agree or disagree with the core elements (i.e., "sensing," "perception," "planning" and "control") described in this document? Please explain why.

A fifth core function should be added: "prediction."

A fifth primary function of prediction will prove critical to safety. It is not sufficient for a vehicle to drive to where the free space is. It must drive to where the free space is going to be when it gets there. There are different methods of prediction suitable for different ODDs that take parts from both perception (e.g., prediction of potential imminent behaviors and changes in behavior based on object type) and planning (e.g., trajectory extrapolation). However, this ANPRM presumes that prediction is part of planning (page 78063), potentially precluding a system design in which prediction is performed in part or in whole as part of the perception function.

Prediction is so critical to safety that it should be treated as a first-class citizen in this list. Moreover, treating prediction as a separate topic for future FMVSS purposes would avoid unnecessarily predetermining whether that function is associated with perception, associated with planning, or an entirely independent vehicle function.

Recommendation: change prediction to be a primary function of: "how the ADS determines the likely future location of relevant objects ('prediction')"

Reference: ANSI/UL 4600 section 8.7 Prediction. Freely available voting draft version available here: http://ul4600.com

Question 8:

At this early point in the development of ADS, how should NHTSA determine whether regulation is actually needed versus theoretically desirable? Can it be done effectively at this early stage and would it yield a safety outcome outweighing the associated risk of delaying or distorting paths of technological development in ways that might result in forgone safety benefits and/or increased costs?

Regardless of the form (e.g., voluntary guidance vs. regulation), progress can be made by emphasizing transparency along the other key points in this response: industry standards, safety first, human operators, and safety cases. There is no need nor any reason to wait in making progress in these areas.

Our comments recommend approaches that adopt industry-generated content and emphasize reporting real-world outcomes so as to perform continuous improvement. This can avoid locking in potentially immature specific technology or practices.

Question 9:

If NHTSA were to develop standards before an ADS-equipped vehicle or an ADS that the Agency could test is widely available, how could NHTSA validate the appropriateness of its standards? How would such a standard impact future ADS development and design? How would such standards be consistent with NHTSA's legal obligations?

NHTSA should emphasize increased adoption of existing and proposed industry-created standards. This includes the following existing normative standards issued by accredited standard development organizations: ISO 26262, ISO 21448 and ANSI/UL 4600.

Question 10:

Which safety standards would be considered the most effective as improving safety and consumer confidence and should therefore be given priority over other possible standards? What about other administrative mechanisms available to NHTSA?

The listed standards of ISO 26262, ISO 21448 and ANSI/UL 4600 are an appropriate set of standards created using industry consensus processes and issued by accredited standards development organizations (SDOs). Additional consideration should be given to a comparable security standard relevant to safety (for example, to mitigate the risk of injection of malicious software that causes unsafe vehicle behavior). Adoption of future SDO-issued standards should not be precluded.

Question 12:

What types and quanta of evidence would be necessary for reliable demonstrations of the level of performance achieved for the core elements of ADS safety performance?

This is still an immature area of struggle for many (if not all) developers of ADS technology. While it is essential for manufacturers to hypothesize and track performance goals of core elements as an essential part of their safety case, it is premature to standardize such metrics other than to say that manufacturers should be attempting to define leading metrics that have predictive value for their specific system and ODD based on their particular safety case.

On the other hand, it is reasonable to standardize lagging metrics for safety outcomes. The plethora of sensors and computational power on ADS-equipped vehicles makes it viable to create much more nuanced lagging metrics such as “near-miss” incidents and traffic regulation infractions or near-miss infractions. Such metric reporting must be non-punitive to be viable. However, it could include data not only for an ADS equipped vehicle, but also for other traffic participants to help monitor baseline norms (suitably anonymized) for comparison with ADS equipped vehicle performance. This data collection capability has the potential to dramatically improve the ability of the industry to understand and improve safety.

Additionally, it is reasonable to encourage corrective action when manufacturer-defined leading safety metrics are violated. (It is important to note that violation of a leading metric is not necessarily a defect. Rather, it is a vital feedback mechanism in a continuous improvement process. Such metric violations are primarily a defect-type issue if they persist long enough that they will plausibly result in reasonably avoidable loss events. And again, this should be a non-punitive process to be viable.)

Question 13:

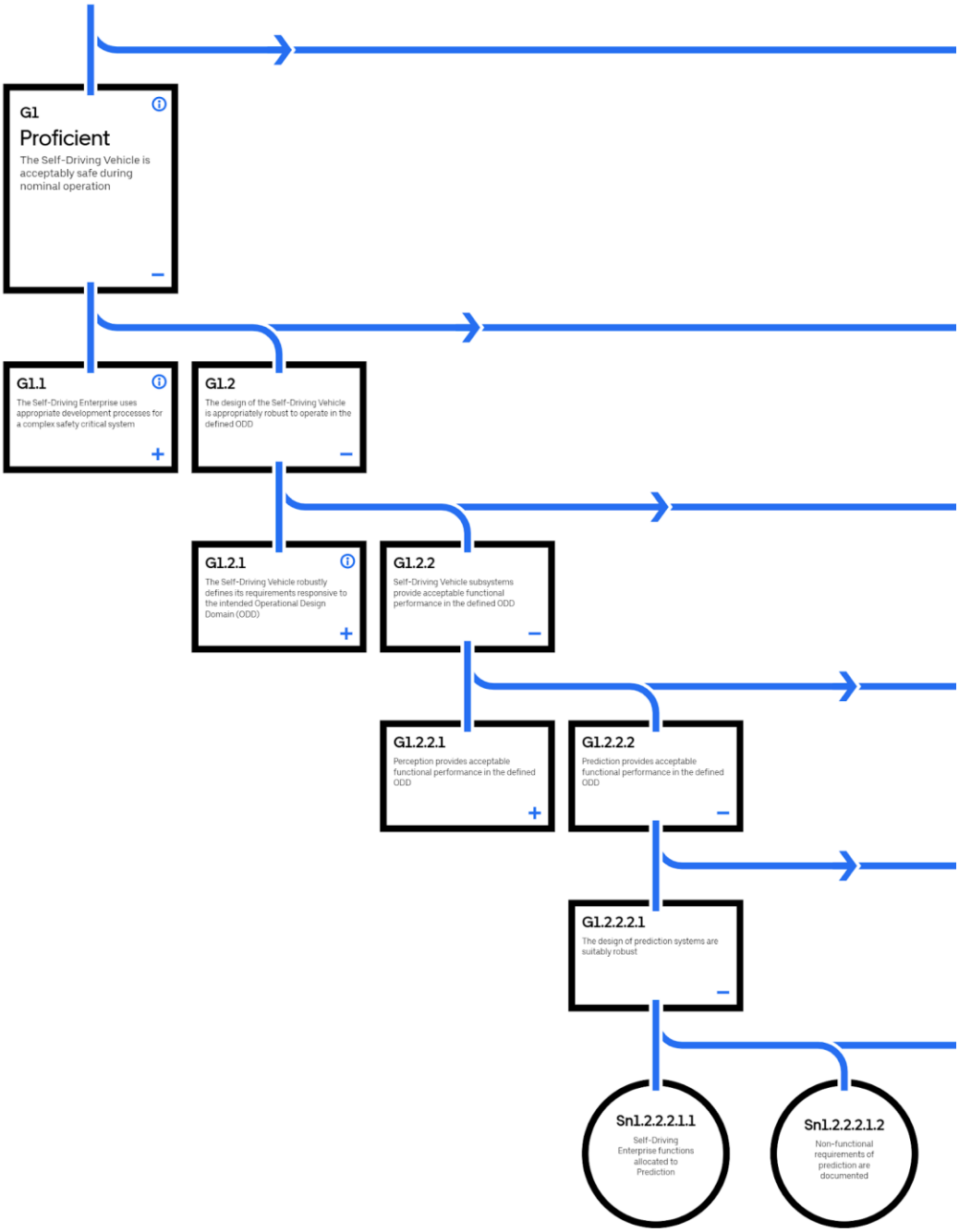
What types and amount of argumentation would be necessary for reliable and persuasive demonstrations of the level of performance achieved for the core functions of ADS safety performance?

Any such demonstration would need to encompass both the necessary level of performance as required by the specific ADS safety case and the validity of the test that the manufacturer says demonstrates an acceptable level of performance.

On the topic of transparency, disclosure of a high level safety case for scrutiny by the public could go a long way to both improving public trust in ADS technology and encouraging reasonable argumentation practices. Such disclosure need not disclose highly confidential design information. As an example, Uber ATG has publicly released two generations of their full safety case framework, with the second generation said to be generally compatible with the ANSI/UL 4600 standard. This is a technically substantive document with hundreds if not thousands of information points. I am not aware of a credible reason to keep high level safety case information secret, and the Uber ATG safety case is an example that such secrecy is indeed not required.

A screen shot of one small piece is included below to illustrate the technical substance of this disclosure (source: https://uberatgresources.com/safetycase/gsn on 22 January 2021):

Our Self-Driving Vehicles are acceptably safe to operate on public roads



B. Question About NHTSA Research

Question 14

What additional research would best support the creation of a safety framework? In what sequence should the additional research be conducted and why? What tools are necessary to perform such research?

1. NHTSA should conduct research into how to best characterize baseline vehicle operational risk at a fine enough granularity and with enough descriptive rigor that the results can be used by manufacturers to know whether they are, indeed, at least as safe as a human driver in comparable operational conditions (or whatever a suitable safety expectation might be). Such data likely exists at least to a degree, but a methodical and uniform approach to expressing a comparison of ADS-equipped vehicle safety to human driver performance using that data would support transparency. Such research should include definition of lagging Safety Performance Indicator metrics that are amenable to apples-to-apples (human to ADS driver) safety comparisons in reasonably specific ODDs or ODD subsets. ADS equipped vehicle performance might not be comparable across vehicle designs due to operation in different ODDs, and any metrics approach should support that possibility.

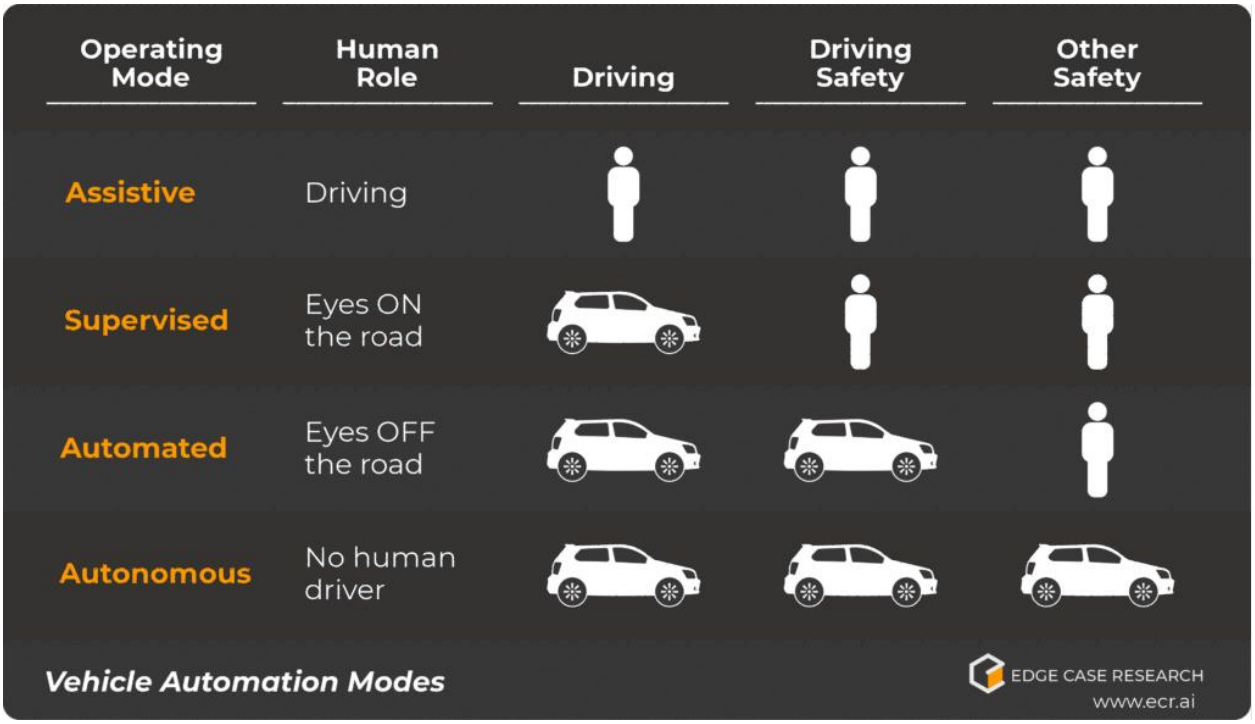
2. NHTSA should support research in defining and evaluating leading metrics, and in particular Safety Performance Indicator metrics tied to safety cases as encompassed by ANSI/UL 4600. While such metrics can be defined now, the industry will benefit from more generalized approaches to both defining and evaluating the predictive power of such leading metrics.

3. NHTSA should conduct research into the practical effectiveness of driver monitoring systems as they relate to any tasks related to human operators. This especially includes the driver fallback function in SAE Level 3 vehicles, but also more generally effectiveness at mitigating automation complacency in SAE Level 2 vehicles.

4. NHTSA should ensure that safety considerations during and after fallback and/or minimum risk maneuvers are addressed. It is not always the case that an ODD exit can be predicted sufficiently far in advance to ensure the vehicle always stays in its intended ODD. There will be cases in which vehicles are forcibly ejected from their ODD without sufficient warning to come to a safe state before departing the ODD (simple example: unforeseen rain squall hits a sunny-day-only vehicle mid-trip). Just because a reasonably foreseeable event has been declared outside the ODD by the vehicle design team does not mean it is OK for such an event to subject road users to unreasonable risk. In general, discussions about how to describe and handle ODD departures as well as talk about safety about what happens during and after such events are an undeveloped area, and worthy of further study research. The topic of how to assess and evaluate risk inherent in ODD departures, fault mitigation situations, and other similar topics should be specifically considered when contemplating the overall system safety of highly automated vehicles. This includes addressing the risk inherent in Minimal Risk Condition (MRC) states, since “minimal risk” might still be unacceptably dangerous in some situations.

5. NHTSA should sponsor research to find more effective ways to communicate roles and responsibilities to vehicle drivers. Such research is required in the face of so much well publicized misuse of less-than-full automation. It could also inform naming of automation features in a more consumer-friendly manner to promote safe and responsible driver behavior.

As part of this, NHTSA should consider adopting the Vehicle Automation Mode descriptive framework depicted below:



(Source: https://edge-case-research.com/project/a-users-guide-to-vehicle-automation-modes/ Note: this figure is released for public redistribution via a Creative Commons BY 4.0 open use license per https://creativecommons.org/licenses/by/4.0/)

This framework emphasizes the driver responsibility, and could go a long way to clearing up the ubiquitous confusion that ordinary drivers have about their responsibility in such vehicles – especially for SAE Level 2 and SAE Level 3 vehicles, as well as vehicles said to be “Level 2+” even though no such SAE Level actually exists.

A useful NHTSA research topic would be to compare a driver-centric representation such as this versus the SAE J3016 engineering-centric approach when communicating to non-technical drivers. Does such an approach result in an improvement of driver understanding and, potentially, performance of driver role, especially in less-than-fully-automated vehicles?

General comments:

It is good to see NHTSA endorsing the findings of the Transportation Research Board that "Careful adherence to process standards can enhance the safety of finished motor vehicles substantially." (ANPRM page 78065). We wholeheartedly agree with this finding that is widely embraced across other safety critical computer-based system development industries, which use process standards as a fundamental basis for achieving safety.

Regarding the metrics discussions, we believe that while it is reasonable to define broad lagging metrics that apply to all vehicles, universally applicable leading metrics will prove more elusive. That is because, at least for the time being, the technology used, the technical architecture, and the corresponding safety cases will vary significantly across different companies and vehicles. Thus, a productive strategy should include both working on defining lagging metrics to provide a level playing field as well as encouraging manufacturers to define their own system-specific leading metrics to inform continuous safety improvement.

Nothing in these comments should be characterized as approving the operation of unreasonably risky vehicles on public roads. Rather, the feedback mechanisms proposed are a way to implement continuous improvement of vehicles that manufacturers should release to deployment with a good faith belief in acceptable safety, but with the realization that deploying currently-immature ADS technology will inevitably lead to unexpected safety-relevant outcomes that can and should be improved.

Nothing in these comments should be construed as advocacy for a particular means of incentivization and/or enforcement. All comments should be implemented with the least enforcement pressure that is viable to achieve the objectives. However, an escalation mechanism for unreasonably risky vehicle designs and operations must be both present and credible to ensure public safety.

For purposes of transparency, I am co-founder of Edge Case Research, which is involved in providing tools and services to stakeholders for autonomous vehicle development and deployment.

Additional comment regarding the use of the term "ADS"

pp. 78058 – 78059. Section I. Current text: "An ADS is the hardware and software that are, collectively, capable of performing the entire dynamic driving task on a sustained basis, regardless of whether it is limited to a specific operational design domain (ODD). In less technical terms, an ADS maintains the control and driving functions within the situations that the system is designed to operate in."

While this is a definition of an ADS in line with SAE J3016, that document itself explicitly puts some safety critical aspects out of scope if unrelated to the Dynamic Driving Task (DDT). An

example safety concern out of scope for J3016 is post-crash safety behaviors, even though that is included in the NHTSA VSSA guidance.

NHTSA should consider all aspects of highly automated vehicle safety rather than just the safety of the ADS performance of the Dynamic Driving Task (DDT) per the SAE J3016 definition when considering rule making. While this is already happening to a degree (for example the VSSA topic just mentioned), NHTSA should be more explicit that they are dealing with holistic highly automated vehicle system safety (for example, as considered in the scope of ANSI/UL 4600) rather than just SAE J3016-definition ADS safety as indicated in the ANPRM introduction. Specifically included should be both the scope of automation that supports safety beyond driving functions and the safety implications of human/computer interactions.

Suggestion: use the term Highly Automated Vehicle Safety (HAV Safety) rather than ADS Safety in future documents to avoid the chance for confusion about the scope of the discussion. This term has been previously used by NHTSA, and seems more encompassing of the mission at hand. An alterative term is Vehicle Automation Safety, which avoids confusion with any specific “automated” operation modes.