



October 17, 2025

**RE:** Docket Number DOT-OST-2025-1029  
U.S. Department of Transportation, Office of the Assistant Secretary for Research and Technology

**Respondent:**

The University of Michigan Transportation Research Institute (UMTRI)

**Contact Information:**

Henry Liu, Ph.D.  
Director, University of Michigan Transportation Research Institute & Mcity  
Professor, Civil and Environmental Engineering  
Bruce D. Greenshields Collegiate Professor of Engineering  
2901 Baxter Road  
Ann Arbor, Michigan 48109  
(734) 764-4354  
[henryliu@umich.edu](mailto:henryliu@umich.edu)

**Introduction to UMTRI and Mcity**

The University of Michigan Transportation Research Institute (UMTRI) is one of the world's premier academic transportation research institutes, dedicated to advancing safety and mobility through interdisciplinary research and public-private partnership. Mcity, founded by UMTRI as the world's first dedicated proving ground for connected and automated vehicles, brings together controlled-environment testing, simulation at scale, and instrumented urban operations. UMTRI offers a uniquely integrated stack for AV safety and impact evaluation. Mcity 2.0 is a remote-access, simulation-in-the-loop testbed that allows researchers and operators to execute scenario-based evaluations that trace from city-scale digital twins to the proving ground's configurable streetscapes and back again, with identical metrics and stimuli. UMTRI's open-source city-scale scenario generation and simulation capabilities support naturalistic and adversarial driving environment modeling, enabling the synthesis of rare but consequential events grounded in real behavior. The Ann Arbor Connected Environment 2.0 has C-V2X roadside units and perception sensors at approximately 75 intersections, paired with an edge-cloud analytics platform capable of near-miss detection and; this platform has already demonstrated its value by winning the USDOT's Intersection Safety Challenge and informing local safety interventions. For measurement, UMTRI developed the Ground Truth Trip Recorder (GTTR) to generate comparable on-road evidence across human drivers, ADS, and hybrid operation. Mcity's Behavioral Safety Assessment Program formalizes a two-part approach, basic behavioral competency and Driving Intelligence Test, to evaluate core rule-of-road behaviors and complex multi-agent interactions. Beyond technology, we maintain deep partnerships with state and local agencies, industry

University of Michigan Transportation Research Institute  
2901 Baxter Road, Ann Arbor, MI 48109, USA

operators, and standards bodies, which allows us to translate research into policy-ready guidance and deployable tools. As a public institution, our mission is to generate evidence that serves the broader public interest. We stand ready to help USDOT build the national infrastructure for AV safety evidence that will provide the clarity, comparability, and confidence that communities deserve.

### **Executive Overview of UMTRI's Response to the RFI**

UMTRI's central thesis for this response is that the nation needs a transparent, evidence-based, and scalable approach to assess and assure AV safety in real-world context. This approach must be anchored in standardized data schemas for apples-to-apples comparisons; representative human-driver baselines; and rigorous, causal evaluation methods that potentially include randomized controlled trials. We recommend the following strategies:

- Establish a National Corner-Case Registry (NCCR) centered on near-misses, not just crashes, to overcome the “curse of rarity.” Use standardized, privacy-preserving scenario fingerprints from both vehicle and infrastructure sensors so evidence can be generalized, reproduced, and stress-tested nationwide.
- Stand up a city-scale National Digital Proving Ground (city-scale digital twin) to evaluate full AV stacks at scale and run randomized controlled trials (RCTs) that benchmark AVs against human drivers before field pilots.
- Launch a rigorous third-party behavioral safety testing framework that is ODD-specific and probabilistic, reports uncertainty with standardized metrics, and integrates with public research hubs for transparent, apples-to-apples comparisons across fleets and cities.
- Modernize Event Data Recorder for AV safety monitoring and build a library of validated crash surrogates
- Advance Accessible-by-Design AVs with independent wheelchair securement and validated occupant protection

### **Suggestion #1: Establish a National Corner-Case Registry to Accelerate AV Development and Validation (addresses RFI Questions 1 and 2)**

Today's core challenge in AV safety is statistical: the events that most threaten safety and public trust are rare, high-dimensional, and context-dependent, which makes naive accumulation of miles inefficient and inconclusive. Our research formalized this problem as the AV “curse of rarity,” [1] articulating why safety-critical events are too infrequent and too diverse to be captured reliably without targeted strategies, and outlining approaches to turn rare, heterogeneous incidents into learnable scenario families.

We recommend that the U.S. Department of Transportation establish a National Corner-Case Registry (NCCR) expressly designed to overcome the “curse of rarity” in AV

safety. A national Corner-Case Registry would operationalize this by creating uniform, privacy-preserving submissions that assign each incident a reproducible “scenario fingerprint”, combining semantic labels, roadway and environmental context, and standardized surrogate-safety time series, so that evidence from one jurisdiction can be generalized, reproduced, and stress-tested elsewhere.

We note that the corner cases here are not necessarily crashes; in fact, the majority should be near-misses, not collisions. These behavioral precursors to harm create far richer, faster signals for learning and validation than sparse crash counts or heterogeneous disengagement tallies. Current regimes such as California’s disengagement reporting provide little detail about what actually transpired. A near-miss-centric registry remedies these shortcomings by standardizing what is measured, how context is encoded, and how evidence is curated across cities and operators.

The NCCR could accept submissions from both vehicle-based sensors (perception and telemetry) and infrastructure-based sensors (roadside cameras, lidars, and C-V2X), enabling multi-vantage reconstructions of the same event. Each submission would be converted into a reproducible scenario fingerprint that joins semantic labels (for example, occluded pedestrian emergence, left-turn-across-path, emergency-vehicle approach), geometric and environmental context (signal phase, sight lines, weather/illumination, temporary traffic control), and a short, standardized time series of surrogate-safety measures such as post-encroachment time, time-to-collision, conflict probabilities, hard-braking intensity, and lateral stability, paired with clearly defined exposure denominators for the relevant ODD. A governed intake process would handle deduplication when multiple contributors observe the same phenomenon, maintain provenance and versioning, and preserve an auditable chain of custody suitable for third-party analysis.

Funded by the US DOT, the Ann Arbor Connected Environment (AACE) [2], combining roadside perception, C-V2X, and edge-cloud near-miss detection, can seed an initial reference node and help publish open schemas, APIs, measurement definitions, and interoperability tests. Our infrastructure-integrated near-miss analytics and validation methodology were recognized by the U.S. DOT’s Intersection Safety Challenge, where the University of Michigan team was selected as a Tier-1 winner in Stage 1B, underscoring the technical readiness and national relevance of our approach. As a Registry node, AACE can ingest standardized event traces, publish de-identified scenario bundles with evidence-grade metadata, and feed an open, replicable pipeline for scenario mining and re-creation across cities and proving grounds.

Centering the NCCR on near-misses and calibrated surrogates, rather than crashes or disengagement counts, directly addresses **RFI Question 1**’s call for comparable, confidence-aware data frameworks and **Question 2**’s need for a systematic method to identify, characterize, and track edge cases over time, accelerating AV development and safety validation without waiting for crashes to happen.

**Suggestion #2: Establish a City-Scale National Digital Proving Ground to Evaluate AV Impacts on the Transportation System (addresses RFI Question 4)**

The nation needs a safe, repeatable, and policy-relevant way to measure how automated vehicles affect the transportation system before and alongside large-scale deployment. Observational comparisons are confounded by shifting demand, policies, and operating design domains; crash outcomes are too rare to provide timely guidance, and single-site pilots cannot easily be generalized. A city-scale digital twin solves this by providing a validated, sensor-faithful, city-scale environment where agencies and operators can run standardized experiments that yield causal, uncertainty-quantified estimates of safety and mobility impacts.

UMTRI and Mcity have laid the technical foundation for such a program through our city-scale, end-to-end, open-source toolchain built around TeraSim [3][4] and our Naturalistic and Adversarial Driving Environment (NDE/NADE) methodology [5][6]. TeraSim synthesizes statistically realistic traffic at metropolitan scale, calibrated to real demand, controls, and infrastructure, including occlusions, sight triangles, work zones, weather, and vulnerable road user flows, so that automated driving stacks encounter the same families of situations they will face in the wild. NADE extends this realism with principled scenario generation that concentrates learning and evaluation on behaviorally meaningful, safety-critical situations while preserving the background statistics that matter. Together, this enables evaluation of full AV stacks “end-to-end,” from perception to planning to control, interacting with human drivers and VRUs, under the same exposure definitions, surrogate safety measures, and network performance metrics that will be used in the field.

To ensure results that the public and policymakers can trust, experiments in the digital twin follow clinical-style rigor, including a randomized controlled trial (RCT) design [7] to benchmark AV safety performance against carefully matched human drivers. In the digital twin, we can randomize routes, time windows, or market shares between AV agents and human-driver controls, pre-register endpoints and analysis plans, stratify by ODD slices, and report uncertainty transparently—thereby de-risking protocols and instrumentation before any field exposure. Once validated in the twin, these same RCT protocols can transition to field tests, if possible.

In short, a city-scale digital twin operationalizes **RFI Question 4**’s call for nationwide, safe, transparent, and equitable evaluation. It delivers a common language of scenarios, metrics, and uncertainty. It evaluates complete AV stacks at city scale. It creates a repeatable pathway, from hypothesis to digital RCT to field deployment, for evidence that communities and policymakers can rely on. UMTRI and Mcity are prepared to stand up the reference implementation, connect it to standardized data and reporting pipelines, and collaborate with USDOT and local agencies to make results comparable across the country.

**Suggestion #3: Establishing a rigorous and trust-building third-party testing framework for automated vehicle safety (addresses RFI Question 5)**

A clear, nationally recognized safety benchmark reduces uncertainty for consumers, investors, insurers, and state regulators, while rewarding companies that invest in genuine safety engineering. For industry, a trusted third-party framework yields resilient safety cases that provide context when incidents occur, demonstrating due diligence through certified behavioral evidence and risk metrics. It also creates a level playing field in which firms can compete on demonstrated performance rather than marketing narratives, and it accelerates market access and interstate scaling by supplying a common yardstick that reduces duplicative, state-by-state negotiations. Importantly, the framework does not prescribe designs or slow iteration; it asks a practical, market-oriented question: how safe is the product in its intended context—quantified with uncertainty?

A national third-party testing program should underscore two principles essential for AI-driven systems. First, safety validation must accommodate probabilistic outputs rather than assume determinism. Modern perception, prediction, and planning are inherently stochastic. Therefore, safety must be expressed and verified as rates and probabilities with confidence bounds. Traditional pass/fail compliance against a checklist cannot capture this reality. A credible program quantifies risk as probabilistic targets. For example, the probability of hazardous conflict per vehicle-mile within a specified ODD uses statistically valid evidence with pre-declared power and uncertainty bounds, binds each certification to the ODD actually tested with composable expansion rules for new geographies or conditions, and supports continuous assurance so that software updates, sensor changes, and model retraining trigger change-impact analysis and re-testing under the same probabilistic criteria.

Second, behavioral safety must be at the center, complementing, not replacing, functional safety (ISO 26262) and SOTIF (ISO 21448). Process conformance is necessary, but the decisive factor for public-road operation is how an AV actually behaves around people and other road users in its declared operating design domain. As with human driver licensing, AVs should be judged on demonstrated behavior in context using nationally standardized performance metrics and surrogate safety measures that correlate with crash risk. These include conflict and near-miss indicators such as time-to-collision distributions, post-encroachment time, gap acceptance, and minimum distance to vulnerable road users; control-quality measures such as speed discipline, acceleration and jerk profiles, lane-keeping deviation, and stopping accuracy at crosswalks and in school or work zones; rules-of-the-road and courtesy behaviors such as yielding, right-of-way compliance, and appropriate responses to emergency vehicles and school buses; and robustness metrics that evaluate stability across weather, lighting, and traffic density and after over-the-air updates.

Mcity’s Behavioral Safety Assessment Program offers a practical blueprint for such third-party, performance-based certification centered on behavioral and probabilistic safety. It comprises two complementary components that operators can adopt now. The first, Basic Behavioral Competency Testing, is a black-box, ODD-specific scenario library that verifies foundational driving behaviors, both routine and safety-critical, using risk-tiered scenarios and standardized metrics (for example, TTC/PET, control quality, and yielding performance). This establishes a behavioral floor analogous to human licensing before large-scale deployment. The second, the Driving Intelligence Test, is a statistical evaluation of overall safety in a naturalistic and adversarial environment that uncovers previously unknown unsafe behaviors and estimates risk as a rate with confidence bounds (for example, crashes or hazardous conflicts per mile) for a defined ODD, accelerating exposure to rare events without bias. Together, these components address both the “known” competencies and the long-tail risks that matter most to the public, while providing the audited, evidence-grade outputs USDOT needs for transparent, hub-integrated reporting.

This suggestion directly answers **RFI Question 5** by specifying how USDOT can enable transparent public understanding of AV impacts. The framework operationalizes Q5’s call for consistent, comparable, and comprehensible reporting that builds public trust.

**Suggestion #4: Modernize Event Data Recorder for AV safety monitoring and build a library of validated crash surrogates**

For decades, event data recorders (EDRs) have been installed on nearly all vehicles sold in the United States, and open crash datasets from NHTSA have advanced understanding of the crash population while accelerating countermeasures that reduce injuries for occupants and vulnerable road users (VRUs). Current EDR regulations (49 CFR Part 563) focus on crash pulses, selected occupant-protection systems, and a handful of pre-crash signals (e.g., brake application and speed). These requirements have yielded valuable insight into driver behavior in the seconds before impact – after all, over 90 percent of crashes involve human-related errors or actions.

Automated driving systems (ADS), however, sense and decide in ways far richer than conventional EDRs capture. Modern AV stacks contain detailed information about the external driving environment and the system’s perception, prediction, planning, and control—along with passenger seating, pose, and restraint status, and nearby VRUs. Extending EDR requirements for AVs is therefore essential to understand crashes and the roles or actions taken by the ADS using the sensors already onboard. Enhanced AV EDRs should record, across pre-crash, crash, and early post-crash windows, the roadway context (location, geometry, traffic-control devices and states), vehicle motion and lane position, the presence and relative motion of other vehicles and VRUs (including emergency vehicles), internal ADS states and control commands, and synchronized camera imagery. Capturing the early post-crash interval is especially important to

diagnose any unexpected behaviors by the ADS and to support transparent, evidence-grade investigations.

In parallel, the nation needs a rigorously validated library of crash surrogates to enable timely safety assessment without waiting for rare crash counts. Surrogates are operational events, defined by thresholds on metrics, that stand in for crashes during on-road evaluation. They occur more frequently than crashes yet remain causally relevant to safety. In medicine, surrogate endpoints are vetted and approved based on rigorous validation studies; in vehicle safety, practice has concentrated on defining metrics, not on establishing validated thresholds. The literature’s validation efforts have often been limited to correlations, in part because few datasets contain both surrogate events and sufficient crash outcomes. SHRP2 is the closest precedent but lacks the comprehensive external sensing needed to validate most surrogates. To close this gap, the field requires a large-scale dataset expressly designed for validation, assembled from high-quality telematics, next-generation naturalistic driving studies with 360-degree sensing, and contributions from ADS manufacturers. That dataset should be paired with epidemiologically sound validation methods and a formal review pathway to designate “validated surrogates” and their thresholds.

UMTRI’s recent USDOT-funded work on the Ground Truth Trip Recorder (GTTR) demonstrates how standardized, portable instrumentation can generate comparable, real-world evidence across driving agents. GTTR retrofits a common sensor and data-acquisition suite onto vehicles to observe safety-related driving performance as they operate on public roads, whether driven by humans, by an ADS, or by hybrid human-automation. This creates a consistent measurement backbone that complements simulation and proving-ground results and feeds both enhanced EDR analysis and surrogate-validation studies. In addition, recent NHTSA-sponsored UMTRI work shows how kinematic safety metrics can be constructed under either simple or more sophisticated modeling assumptions when using surrogate-metric methods [8], and that post-hoc measures derived from evaluation or operational datasets are materially improved when the more advanced models are applied [9].

Suggestion 4 directly advances the RFI’s goals across multiple questions: it strengthens **Q1 (Data standards and integration)** by specifying common, evidence-grade fields for enhanced AV EDRs; it supports **Q2 (Edge-case identification)** by capturing rich pre-crash and near-miss context—across perception, planning, control, and roadway state—so rare hazardous patterns can be discovered, fingerprinted, and tracked; it operationalizes **Q4 (Evidence-based evaluation)** by enabling a rigorously validated library of crash surrogates and a GTTR-style on-road evaluator that produce statistically defensible effect estimates and uncertainty bounds; it fulfills **Q5 (Transparency & public understanding)** by generating auditable records with provenance, timestamps, and standardized metrics suitable for hub publication and apples-to-apples comparisons; and it contributes to **Q6 (Consistent & robust behavior)** by allowing repeatable, cross-

agent measurement of how AVs behave around other road users in varied ODD slices, including post-update regression checks and adverse-condition testing.

**Suggestion #5: Advance Accessible-by-Design AVs with Independent Wheelchair Securement and Validated Occupant Protection (addresses RFI Q3(c))**

We recommend the U.S. DOT to sponsor research to advance accessible-by-design for people with disabilities—especially riders who remain seated in their wheelchairs—by further developing independent wheelchair securement, validating occupant protection systems for wheelchair-seated occupants, and advancing human-machine interfaces (HMIs) that work for diverse users including emergency responders, pedestrians, cyclists, and passengers. Automated driving removes the human driver who, in today’s paratransit workflows, often performs manual wheelchair tiedown; in AV service, riders must be able to board, dock, and ride independently and safely.

UMTRI has built a deep research portfolio to make AVs accessible-by-design for wheelchair users and other riders with disabilities, including the development and evaluation of independent wheelchair securement concepts compatible with Universal Docking Interface Geometry (UDIG) [10] and aligned with RESNA WC19, plus multiple studies on automated tiedown and occupant-restraint systems—from design and volunteer trials to guidance for independent safety in AVs and optimization of frontal-impact protection for wheelchair-seated occupants. UMTRI synthesized this evidence into its Design Guidelines [11], which frames ingress/egress, station sizing, securement, and occupant protection needs for ADS-equipped vehicles.

References:

1. H. X. Liu and S. Feng, “Curse of rarity for autonomous vehicles,” *Nat. Commun.*, vol. 15, no. 1, 2024, doi: 10.1038/s41467-024-49194-0
2. <https://aace.umtri.umich.edu/>
3. H. Sun *et al.*, “TeraSim: Uncovering Unknown Unsafe Events for Autonomous Vehicles through Generative Simulation,” Apr. 02, 2025, *arXiv*: arXiv:2503.03629. doi: 10.48550/arXiv.2503.03629.
4. J. Wang, H. Sun, X. Yan, S. Feng, J. Gao, and H. X. Liu, “TeraSim-World: Worldwide Safety-Critical Data Synthesis for End-to-End Autonomous Driving,” Sept. 17, 2025, *arXiv*: arXiv:2509.13164. doi: 10.48550/arXiv.2509.13164.
5. X. Yan, Z. Zou, S. Feng, H. Zhu, H. Sun, and H. X. Liu, “Learning naturalistic driving environment with statistical realism,” *Nat. Commun.*, vol. 14, no. 1, p. 2037, Apr. 2023, doi: 10.1038/s41467-023-37677-5.
6. S. Feng *et al.*, “Dense reinforcement learning for safety validation of autonomous vehicles,” *Nature*, vol. 615, no. 7953, pp. 620–627, Mar. 2023, doi: 10.1038/s41586-023-05732-2.

7. C. Flannagan, “Safety Benchmarking of Automated Driving Systems: The Challenges of Apples-to-Apples Comparisons”, CCAT Distinguished Lectures, June 4, 2024, at the University of Michigan Transportation Research Institute.
8. LeBlanc, D., Liu, H. X., Flannagan, C. A., Yan, X., Feng, S., and Leslie, A. J. (2021). “Safety metrics for assessing driving performance of automated driving.” Final project report under review for publication by National Highway Traffic Safety Administration.
9. Yan, X., Feng, S., LeBlanc, D. J., Flannagan, C., Liu, H. X., and Svenson, A. L. (2023). “Safety Metrics Assessment Using Logged Vehicle Trajectory Data.” Proc. 27th International Technical Conference on the Enhanced Safety of Vehicles, Paper 23-0224. Yokohama, Japan, April 2023.
10. Klinich, K. D., Boyle, K. J., Manary, M. A., Orton, N. R., Wang, Y., & Hu, J. (2023). Optimizing frontal impact occupant protection systems for passengers seated in wheelchairs. *Accident; Analysis and Prevention*, 192. <https://doi.org/10.1016/J.AAP.2023.107238>
11. Klinich KD, Orton NR, Manary MA (2022) Design Guidelines for Accessible Automated Vehicles: Mobility Focus. UMTRI 2022-8. University of Michigan Transportation Research Institute.