Proc. of the European Safety and Reliability Conference (ESREL 2026)
Edited by José C. Matos, Paulo B. Lourenço, Michael Beer, Daniel Oliveira, Edoardo Patelli, Hélder S. Sousa, Rui A. Silva and Monica Santamaria

doi: 10.3850/ESREL2026061419_esrel26-p27763-cd

Towards safe deployment of autonomous vehicles: responsibilities, hazards, and metrics identification based on STPA methodology

Tingting Cheng[1], Camila Correa-Jullian[1,2], Siddhant Singh[1], Ali Mosleh[1], Jiaqi Ma[2]

[1]*B. John Garrick Institute for the Risk Sciences, University of California, Los Angeles, USA.* tingtingc@ucla.edu; ccorreaj@ucla.edu; mosleh@ucla.edu

[2]*Center of Excellence on New Mobility and Automated Vehicles, University of California, Los Angeles, USA.* jiaqima@ucla.edu

The autonomous vehicle (AV) industry has significantly evolved since the initiation of the AV testing program by the California Department of Motor Vehicles (DMV) in 2012. As commercial deployments for passenger service transport expand to multiple cities, key challenges remain to comprehensively assess the AV fleet's safety performance and their impact on traffic systems. In the U.S., AV developers primarily report limited and siloed incident metrics that offer little insight into how risks may evolve as AV fleets scale. This limits the ability of regulatory entities to make informed decisions about AV integration within complex urban traffic systems. Therefore, this work proposes a model-based methodology to elicit operational safety-related metrics as part of an ongoing effort to develop risk-informed tools enhancing the capacity of local jurisdictions to assess an AV fleet's safety performance. This work leverages System-Theoretic Process Analysis (STPA) to identify and model safety-critical interactions between AV fleets and other road users across deployment stages, aiming to identify hazards arising from unsafe control actions (UCAs) that could result in system-level losses beyond traditional crash metrics, such as delayed emergency response. The STPA-based analysis accounts for interacting entities (e.g., AVs, human-driven vehicles), responsibilities of the entities (e.g., preventive or mitigative controls), control actions (e.g., remote operators send commands to AVs), and feedback mechanisms (e.g., AVs send environmental conditions back to remote operators). UCAs are identified through analysing control actions in the operational context of AV fleets, supported by available operational and incident data, and assessed for public-reporting feasibility. This work presents the functional control and interaction model of a generic AV fleet, a structured catalogue of critical UCAs, associated hazard scenarios, and corresponding performance metrics linked to system safety objectives.

Keywords: Autonomous vehicles, testing and deployment, safety analysis, STPA

1. Introduction

Autonomous vehicle (AV) testing and deployment are expanding across cities with widely varying roadway geometries, traffic compositions, weather patterns, regulatory environments, and social expectations. As SAE Level 4 Automated Driving System (ADS)-equipped fleets (SAE On-Road Automated Vehicle Standards Committee 2021) transition from limited testing to large-scale commercial deployment, the need for systematic, data- and model-driven approaches to evaluate operational safety performance becomes critical. Such approaches are necessary to support informed decisions about when and how fleets may be permitted to advance across deployment milestones (e.g., from pilot to commercial operations). To date, much of the discussion around AV safety performance has either focused on vehicle-level capabilities or high-level aggregate outcomes such as accumulated vehicle-miles-travelled (VMT), incident rates, and disengagements (National Highway Traffic Safety Administration. 2025; California Public Utilities Commission 2025). While these indicators provide useful data points, they are limited in their capacity to both characterize the multi-dimensional risks introduced by AV fleet operations and to accurately represent potential safety benefits. Detailed incident reporting mechanisms, originally developed to support small-scale AV testing, also lack practical pathways for scaling to large, continuously operating fleets.

The San Francisco County Transportation Authority (SFCTA) recently proposed an incremental, performance-based AV permitting framework that links fleet growth and operational complexity to quantitative safety performance thresholds (Castiglione et al. 2025). The framework models cumulative exposure through VMT and defines staged transitions from testing to deployment based on performance across a set of operational safety metrics, including collisions, injuries, unplanned stops, vehicle retrievals, and first-responder obstructions. By explicitly tying operational permissions to measured outcomes, this framework represents a significant step toward developing more transparent, evidence-based risk acceptance discussions at local jurisdiction level. However, the proposed framework, constrained by currently available reporting requirements, largely emphasizes lagging indicators of safety performance, such as collisions and service disruptions due to unplanned stoppages (Bayramova et al. 2023).

Prior work introduced a hazard identification methodology for AV fleet operations that explicitly considered hardware, software, human, organizational, and environmental sources of risk from the standpoint of a fleet operator (Correa-Jullian et al. 2024b; Ma et al. 2024; Correa-Jullian et al. 2024a). The present work reframes the analysis through the lens of local jurisdictions, exploring how operational safety risk may be better characterized, monitored, and compared as AV fleets scale. This work explores how safety metrics beyond crash-related data can inform deployment decisions and operational strategies in a manner consistent with incremental, performance-based permitting. Building on this foundation, the paper proposes a model-driven derivation of safety performance metrics, elicits interaction-based indicators, and provides a risk-informed justification for their relevance to public-facing AV risk acceptance.

2. Methodology

Safety is commonly defined as a state in which risks have been sufficiently managed and the residual risk is considered acceptable (Rausand 2013), i.e., the absence of unreasonable risk (Favarò et al. 2026). Risk, therefore, is assessed through both qualitative and quantitative approaches, and its management involves addressing not only its sources, but also its perception (Bin-Husayn, Abdullah, and Rimmel 2024). This work develops a structured, system-oriented methodology to expand AV safety performance metrics beyond incident-based indicators by identifying risk-relevant interactions and operational hazards at the fleet level.

System-Theoretic Process Analysis is a qualitative hazard identification method based on system control theory (Leveson and Thomas 2018). STPA has been applied to the analysis of human-autonomy interactions (Cheng et al. 2023; Thornberry 2014) as well as to AV fleet management (Correa-Jullian et al. 2024b; 2024a) (Siemens, 2023; Auto AI, 2025), demonstrating robust performance and practical effectiveness. In this work, STPA is applied to model AV fleet operations as complex socio-technical systems characterized by interacting technical, human, organizational, and environmental elements, represented through structured control diagrams. The analysis focuses on AV operations across deployment stages and interactions with human agents, with the goal of identifying hazards arising from unsafe control actions (UCAs). UCAs are identified by examining how control actions may become unsafe when provided incorrectly, omitted, premature, delayed, or applied for an inappropriate duration. This approach is used to construct a catalogue of critical UCAs and associated hazard scenarios linked to a set of candidate performance metrics, traced to system safety objectives.

2.1 *STPA Methodology*

Following standard STPA practice, the analysis begins by defining system-level losses from the perspective of relevant stakeholders, including, in this case, local transportation authorities. Based on these losses, system-level hazards are identified, including but not limited to human injury, property damage, disruption of emergency response, and traffic congestion. Corresponding system-level safety constraints are then specified, defining the conditions that the AV fleet and its supporting organization must satisfy to prevent these hazards.

The next step is to build a hierarchical control structure representing system operation. This structure defines the interacting actors (e.g., AVs, human-driven vehicles), their responsibilities (i.e., safety constraint-related controls, either preventive or mitigative), control actions between entities (e.g., remote operators send commands to AVs), and associated feedback mechanisms (e.g., AVs send environmental conditions back to remote operators). The resulting model represents the system in terms of controllers, controlled processes, feedback, and interactions, organized into interconnected feedback control loops. This structure provides the basis for analysing how inadequate control or feedback may lead to hazardous system behaviour.

2.2 *Proposed STPA model*

The system analysed in this study is a transportation system withing which deployed Level 4 ADS fleets operate. This includes human-driven vehicles (HDVs), VRUs, passengers, and first responders, with oversight and coordination functions provided by the Traffic Management Centre (TMC) and the Transportation Authority (TA). Within the STPA framework, traffic conditions (e.g., safe operation, incidents, accidents of varying severity, and congestion) are treated as the controlled process.

Key stakeholder groups include S-1 HDVs, S-2 VRUs, S-3 Passengers, S-4 Regulatory bodies, S-5 Law enforcement, S-6 First responders, S-7 Fleet operators, and S-8 ADS developers. System-level losses are defined based on the KABCO injury severity scale (National Highway Traffic Safety Administration 2025), supplemented with risk management concepts to capture non-injury impacts relevant to AV deployment. Table 1 summarizes the system-level loss categories considered in this study. System-level hazards are defined as unsafe traffic conditions that arise from interactions among system actors, rather than from isolated component failures. Drawing on transportation theory and AV-related regulatory, industrial, and academic literature, the hazards listed in Table 2 represent emergent states that increase the likelihood of one or more system-level losses. In AV deployment, injury loss (L1) may arise in scenarios involving collisions, tailgating, or delayed emergency response. Asset loss (L2) may result from traffic accidents or physical conflicts that damage vehicles or infrastructure. Mission or service loss (L3) refers to disruptions such as blocked roads, AV service interruptions, or impeded emergency response capabilities. Environmental loss (L4) is generally indirect and becomes prominent in events such as battery fires or incidents involving hazardous material transport. Finally, economic loss (L5) may arise from crashes, delays, or operational disruptions, leading to both direct financial impacts and broader economic consequences.

Table 1 System-level losses in AV deployment.

Loss Types	Description
L1: Injury loss	Harm to human life or health, including fatalities and injuries to passengers, road users, operators, or first responders.
L2: Asset loss	Damage to physical or digital assets such as vehicles, infrastructure, and supporting systems.
L3: Mission or service loss	Failure or degradation of AV or transportation system operations, including service interruptions or delayed emergency response.
L4: Environmental loss	Negative environmental impacts resulting from system-related incidents, including hazardous material releases, battery fires, excessive emissions, or ecological disruption.
L5: Economic loss	Financial consequences of system failure or degradation, including legal costs, operational downtime, reputational harm, and broader economic impacts on stakeholders or the public.

Figure 1 illustrates the high-level STPA control structure for AV deployment. The diagram captures the hierarchical relationships and feedback loops between transportation authorities, traffic management centres, infrastructure, AV fleets, other road users, and external traffic conditions.

3. Case Study Results

Case studies vary along several dimensions, including fleet size, operational design domain (ODD) scope, availability of a human driver or remote assistance, AV capabilities, and allocation of control authority.

Table 2 System-level hazards, traffic features, emergent states, and corresponding losses.

Hazard ID	Emergent State Description	Relevant Losses
H1: Unstable headways	Erratic AV/HDV spacing increases cut-ins, tailgating, and flow instability.	L1, L2, L5
H2: Abrupt slowdowns	Sudden deceleration and shockwaves reduce reaction time and elevate rear-end collision risk.	L1, L2, L5
H3: Congestion / stop-and-go	Persistent congestion from behavioural mismatch or deployment density degrades throughput and increases conflicts.	L3, L5
H4: Spatial blockage	Lane, intersection, or emergency-path obstruction creates bottlenecks and secondary hazards.	L3, L4, L5
H5: Responder access blocked	AV behaviour impedes emergency access or coordination, increasing injury severity and delay impacts.	L1, L3, L5
H6: Interaction breakdown	Misinterpreted intent between AVs, HDVs, or VRUs leads to hesitation or conflict.	L1, L2, L5
H7: Collision risk	Behavioural mismatch, signal ambiguity, or reaction failure culminates in collision scenarios.	L1, L2, L5
H8: Misplaced stops	Improper AV stopping for passenger actions or MRC disrupts traffic flow and raises rear-end and blockage risk.	L1, L2, L3, L5

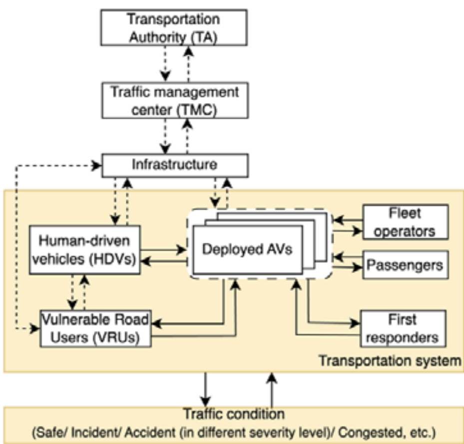

Figure 2. High-level control diagram of transportation environment.

During testing, ADS fleet operations are typically characterized by small fleet sizes, constrained ODDs, and shared control authority between the automated system and an onboard safety driver. In this context, the safety driver serves as a primary fallback controller, mediating passenger and first-responder interactions and compensating for incomplete automation and limited system feedback.

In contrast, in the deployment stage, ADS fleet operations are modelled as a socio-technical control system in which deployed AVs interact with multiple human, organizational, and infrastructural controllers to influence overall traffic conditions. As illustrated in Figure 2, the transportation system is represented as the controlled process, with traffic conditions—ranging from nominal operation to incidents, accidents of varying severity, and congestion—emerging from the combined actions and interactions of system actors. Deployed AVs operate without onboard safety drivers and therefore act as primary controllers within the traffic stream, exchanging bidirectional interactions with human-driven vehicles, vulnerable road users, passengers, fleet operators, and first responders. Oversight and coordination functions are provided at higher levels of the control hierarchy by the Traffic Management Centre and the Transportation Authority, which influence system behaviour through policies, operational constraints, and infrastructure-level controls.

In this driverless deployment context, higher VMT increases system exposure, enabling observation of low-frequency, high-consequence outcomes that are not discernible during limited testing. Deployed AVs operate within expanded operational design domains that may span multiple cities and include nighttime operation, adverse weather, complex intersections, and work zones (Automated Vehicle Safety Consortium 2022; Applied Intuition 2024). Control authority and fallback responsibility shift from human takeover to remote assistance and precisely defined minimum risk manoeuvres, requiring reliable feedback on vehicle state, environment, and traffic conditions (Automated Vehicle Safety Consortium 2023; Correa-Jullian et al. 2025). Operational objectives extend beyond functional safety validation to include service availability and user experience, while maintaining system-level safety constraints. Passenger emergency stop requests must be interpreted and managed autonomously, and AVs must coordinate with first responders during incidents without reliance on a safety driver (Automated Vehicle Safety Consortium 2025). Within this framework, inadequate, incorrect, delayed, or missing control or feedback may give rise to unsafe control actions and hazardous traffic states. Due to space constraints, the main text focuses on one interaction domain particularly relevant for deployment-stage risk: VRU interactions

3.1. Deployed AVs - VRUs

This subsection examines AV interactions with pedestrians, cyclists, and e-scooter users, which are frequent sources of operational complexity and safety risk in urban deployments.

Interactions between deployed AVs and vulnerable road users (VRUs) represent a critical safety-relevant use case due to their high frequency, behavioural variability, and limited tolerance for control or perception errors. From an STPA perspective, VRU interactions involve asymmetric vulnerability and limited coordination mechanisms, placing greater responsibility on the AV to maintain safe control under uncertainty. In deployment, this burden is amplified by the absence of onboard human supervision, expanded operational design domains, and the need for the AV to autonomously interpret intent, negotiate right-of-way, and execute timely fallback actions.

Within the control structure, deployed AVs function as primary controllers influencing VRU safety outcomes through perception, speed regulation, yielding behaviour, and trajectory planning. VRUs, in turn, provide implicit feedback through motion cues, hesitation, evasive manoeuvres, or compliance with traffic rules. Unsafe control actions arise when the AV fails to correctly perceive VRUs, applies control actions too late or with inappropriate magnitude, or does not adequately adapt behaviour to context-specific constraints such as dense multimodal zones, work zones, or time-dependent rules. These failures can lead to hazardous traffic states involving interaction breakdowns or elevated collision risk, even in the absence of direct contact.

3.1.2. *Responsibilities of the deployed AVs*

The AV responsibilities during interactions with VRUs are defined as follows across different deployment stages. These responsibilities serve as the basis for identifying unsafe control actions in both testing and deployment contexts.

- **Pedestrian Interactions:** Detect pedestrians at marked and unmarked crosswalks; Yield and stop when pedestrians are entering or occupying the crosswalk; In the testing stage, provide timely alerts to the safety driver to enable intervention when required.
- **Cyclist Interactions:** Maintain safe lateral distance during overtaking or co-travel; Adjust speed to avoid collision or intimidation, particularly near bike lanes or shared roadways; In the testing stage, rely on safety driver oversight in atypical cyclist behaviours.
- **E-scooter Interactions:** Predict dynamic, erratic movements and avoid unsafe proximity in shared zones; Reduce speed or reroute to maintain safety buffer in congested or multi-modal zones; In the testing stage, defer aggressive rerouting or priority decisions to the safety driver in ambiguous cases.

3.1.2. *UCAs and safety metrics*

Table 3 summarizes the identified UCAs, associated hazards, and candidate safety metrics. categorized as common across testing and deployment (C), deployment-specific (D), or testing-specific (T). Each UCA is linked to anticipated VRU responses, relevant system-level hazards (primarily interaction breakdowns and collision risk), and candidate safety performance metrics that can be operationalized using fleet data.

Metrics may apply to both stages or be specific to the presence or absence of a safety driver. Note that common UCAs reflect failures in perception, yielding, speed control, or interaction negotiation that persist across deployment stages. Deployment-specific UCAs capture challenges arising from increased autonomy, expanded ODDs, and the absence of onboard human supervision. Testing-specific UCAs highlight risks associated with over-reliance on safety drivers and communication breakdowns between the AV and the human operator.

4. Discussion

The proposed VRU-related metrics are derived from unsafe control actions (UCAs) and map directly to operational responsibilities within deployed ADS fleets. Failures in perception, yielding, and speed regulation correspond primarily to ADS-level control, while stopping behaviour, work-zone interpretation, and compliance with time-dependent rules reflect fleet-level configuration and ODD governance. Testing-specific UCAs further illustrate how safety drivers can mask latent control deficiencies, reinforcing the need for metrics that remain valid once human supervision is removed. In this analysis, VRU behaviour is represented implicitly through observable feedback (e.g., hesitation or trajectory changes) rather than as an explicitly modelled agent. This abstraction enables tractable metric derivation across fleets and jurisdictions but does not yet capture jurisdiction-specific interaction norms, such as local yielding practices, which may influence baseline behaviour and warrant future refinement.

Consistent with a system-theoretic framing, crashes are treated as end states rather than primary indicators of control adequacy. UCAs reflect violations of safety constraints, and the associated metrics capture early signs of control degradation. Measures such as time-to-collision, crosswalk yield rate, lateral passing distance, and speed differential in multimodal zones therefore serve as leading indicators of hazardous proximity and interaction breakdown. Given practical constraints on data access and disclosure, these metrics are evaluated in terms of feasibility, availability, analytical value, and reporting burden.

4.1 *Metric Evaluation*

Each candidate metric was assessed using a five-factor scale, with ratings of High (H), Medium (M), or Low (L) for availability, granularity, feasibility of external sharing, safety relevance, and cost to collect. Based on these assessments, metrics were grouped into three reporting tiers. Tier 1 metrics are safety-relevant, low-burden, and suitable for standardized public reporting in aggregated form. Tier 2 metrics provide strong analytical value but require controlled sharing due to sensitivity or contextual dependence. Tier 3 metrics are primarily diagnostic and are best evaluated through audits or regulator–operator review mechanisms.

For instance, Work-zone path deviation and work-zone VRU conflict metrics illustrate Tier 2 cases: both are strongly safety-relevant and offer high diagnostic value for deployment-stage risk, but depend on context-rich data (e.g., temporary signage, geofences, planner behaviour) that complicates external disclosure without aggregation or safeguards. In contrast, school-zone compliance and blocked bike-lane incident rate exemplify Tier 1 metrics. These rely on well-defined map features, vehicle state data, and time windows, impose low incremental collection burden, and can be reported as aggregated rates without exposing proprietary system logic, making them suitable candidates for public reporting without exposing proprietary system behaviour. While the proposed metrics are not intended as standalone predictors of crash risk, they may provide a structured basis for early detection of control degradation and targeted safety assurance as AV fleets scale.

5. Concluding remarks

This paper shows how STPA can be extended from hazard identification to the derivation of operational safety performance metrics for ADS fleets. By modelling AV deployment as a socio-technical control system and tracing UCAs to measurable indicators, the approach connects system-theoretic safety analysis with performance-based AV oversight. A deployment-stage case study focused on VRU interactions demonstrates that many safety-critical conditions arise from interaction breakdowns and mistimed or inappropriate control actions rather than from collisions alone.

Table 3: Identified UCAs and safety metrics for AV-VRU interactions.

UCA-AV	UCAs	VRU feedback	Hazards	Metrics
UCA-AV-VRU-C1	Fails to detect pedestrian at crosswalk *(or fails to alert safety driver in testing)*	Pedestrian proceeds assuming right-of-way	H6, H7	Time-to-Collision (TTC); Crosswalk Yield Rate; Driver Alert Latency *(testing)*
UCA-AV-VRU-C2	Detects pedestrian but fails to yield or stop *(or delays driver takeover in testing)*	Pedestrian stops abruptly or swerves	H6, H7	Stop Distance; Pedestrian Conflict Rate; Time to Driver Intervention *(testing)*
UCA-AV-VRU-C3	Fails to maintain safe lateral distance during overtaking	Cyclist swerves or brakes defensively	H6, H7	Lateral Passing Distance; Cyclist Near-Miss Frequency
UCA-AV-VRU-C4	Maintains excessive speed during overtaking	VRU evasive response due to close pass or wind blast	H6, H7	Speed Differential; Post-Overtake Swerve Events; Driver Awareness Confirmation *(testing)*
UCA-AV-VRU-C5	Fails to anticipate erratic movement in shared zones	E-scooter veers or brakes suddenly	H6, H7	Trajectory Deviation; Sudden Deceleration Events
UCA-AV-VRU-C6	Does not reduce speed in dense VRU traffic *(or fails to cue driver in testing)*	VRUs adjust path or speed to avoid AV	H6	Speed in Multimodal Zones; VRU Avoidance Incidents
UCA-AV-VRU-D1	Misinterprets temporary work-zone detours	Unexpected AV path shifts or blocked crossings	H6, H7	Work-Zone Path Deviation; Cone/Sign Misread Rate; Work-Zone VRU Conflicts
UCA-AV-VRU-D2	Fails to comply with time-dependent VRU rules (e.g., school zones)	VRUs expect yielding or speed reduction not provided	H6, H7	School-Zone Compliance; Time-Window Yield Rate; Time-Stratified VRU Conflicts
UCA-AV-VRU-D3	Stops in bike lanes	Cyclists merge into traffic or stop abruptly	H7	Bike-Lane Occupancy Duration; Blocked Bike-Lane Incident Rate
UCA-AV-VRU-T1	*Over-relies on safety driver for routine VRU interactions*	*Delayed AV response; VRU confusion*	*H6*	*Driver Override Frequency; VRU Interaction Ratio*
UCA-AV-VRU-T2	*Miscommunicates VRU intent to driver (e.g., incorrect visualization or alert)*	*Inappropriate driver manoeuvre; VRU startled*	*H6*	*Alert Accuracy Rate; Driver Misinterpretation Incidents (testing)*

To support practical adoption, candidate metrics were evaluated for availability, analytical value, feasibility of external sharing, and reporting burden, and organized into reporting tiers aligned with realistic disclosure constraints. Together, these results illustrate how STPA-derived metrics can support incremental, exposure-normalized, performance-based AV permitting across fleets and operational design domains.

Acknowledgement

This material is based upon work supported by the Federal Highway Administration under Agreement No. 693JJ32350027. Any opinions, findings, and conclusions or recommendations expressed in this publication are those of the Author(s) and do not necessarily reflect the view of the Federal Highway Administration.

References

Applied Intuition. 2024. "Defining ODD Boundaries in 2 Phases." 2024. https://www.appliedintuition.com/blog/odd-taxonomy-a-2-phase-approach-to-defining-testing-boundaries.

Automated Vehicle Safety Consortium. 2022. "AVSC

Best Practice for Interactions Between ADS-DVs and Vulnerable Road Users (VRUs)." SAE Industry Technologies Consortia.
———. 2023. "AVSC Best Practice for ADS Remote Assistance Use Case." *SAE Industry Technologies Consortia*.
———. 2025. "AVSC Best Practice for Automated Driving System-Dedicated Vehicle (ADS-DV) Post-Crash Behaviors and Interaction." SAE Industry Technologies Consortia.
Bayramova, Aya, David J. Edwards, Chris Roberts, and Iain Rillie. 2023. "Constructs of Leading Indicators: A Synthesis of Safety Literature." *Journal of Safety Research* 85 (June):469–84. https://doi.org/10.1016/J.JSR.2023.04.015.
Bin-Husayn, Ali Ramadhan, Maizatulakma Abdullah, and Gunnar Rimmel. 2024. "The Relationship between Risk Perception and Risk Management: A Systematic Literature Review." *Journal of Risk Research* 27 (10): 1290–1307. https://doi.org/10.1080/13669877.2024.2447258.
California Public Utilities Commission. 2025. "Autonomous Vehicle Passenger Service Programs." 2025. https://www.cpuc.ca.gov/regulatory-services/licensing/transportation-licensing-and-analysis-branch/autonomous-vehicle-programs.
Castiglione, Joe, Drew Cooper, Jean Paul Velez, Stephen Chun, and Abe Bingham. 2025. "Conceptual Safety-Focused AV Permitting Framework." https://www.sfcta.org/reports/conceptual-safety-focused-av-permitting-framework.
Cheng, Tingting, Ingrid Bouwer Utne, Bing Wu, and Qing Wu. 2023. "A Novel System-Theoretic Approach for Human-System Collaboration Safety: Case Studies on Two Degrees of Autonomy for Autonomous Ships." *Reliability Engineering & System Safety* 237 (September):109388. https://doi.org/10.1016/j.ress.2023.109388.
Correa-Jullian, Camila, Marilia A. Ramos, Ali Mosleh, and Jiaqi Ma. 2024a. "An STPA-Based Analysis of Automated Driving Systems Fleet Maintenance Activities." In *2024 Annual Reliability and Maintainability Symposium (RAMS)*, 1–6. IEEE. https://doi.org/10.1109/RAMS51492.2024.10457750.
Correa-Jullian, Camila, Marilia Ramos, Ali Mosleh, and Jiaqi Ma. 2024b. "Operational Safety Hazard Identification Methodology for Automated Driving Systems Fleets." *Proceedings of the Institution of Mechanical Engineers, Part O: Journal of Risk and Reliability*, February. https://doi.org/10.1177/1748006X241233863.
———. 2025. "Human-Autonomy Teams: The Case of Remote Operators and Automated Driving Systems." In *Proceedings of the 35th European Safety and Reliability & the 33rd Society for Risk Analysis Europe Conference*, edited by Eirik Bjorheim Abrahamsen, Terje Aven, Frederic Bouder, Roger Flage, and Marja Ylönen, 1957–64. Singapore: Research Publishing. https://doi.org/10.3850/978-981-94-3281-3_ESREL-SRA-E2025-P3614-cd.
Favarò, Francesca Margherita, Laura Fraade-Blanar, Scott Schnelle, Trent Victor, Mauricio Peña, Johan Engström, John Scanlon, Kris Kusano, and Dan Smith. 2026. "Building a Credible Case for Safety: Approach Proposal for Automated Driving Systems." *Journal of Safety Research* 96 (4): 181–96. https://doi.org/10.1016/j.jsr.2025.10.019.
Leveson, Nancy, and John Thomas. 2018. "STPA Handbook." 2018. https://psas.scripts.mit.edu/home/get_file.php?name=STPA_handbook.pdf.
Ma, Jiaqi, Camila Correa-Jullian, Marilia Ramos, and Xin Xia. 2024. "Risk Assessment for Remotely Operation of Level 4 Automated Driving Systems in Mobility as a Service Transport." UC Office of the President: University of California Institute of Transportation Studies. https://doi.org/https://doi.org/10.7922/G23N21QC.
National Highway Traffic Safety Administration. 2025. "Standing General Order, ADS Incident Report Data." 2025. https://www.nhtsa.gov/laws-regulations/standing-general-order-crash-reporting.
National Highway Traffic Safety Administration. 2025. "MMUCC Guideline Model Minimum Uniform Crash Criteria, 6th Edition." *Report No. DOT HS 813 525a*.
Rausand, Marvin. 2013. *Risk Assessment: Theory, Methods, and Applications*. Vol. 115. John Wiley & Sons. https://books.google.com/books/about/Risk_Assessment.html?id=9EHeLmbUVh8C.
SAE On-Road Automated Vehicle Standards Committee. 2021. "Taxonomy and Definitions for Terms Related to Driving Automation Systems for On-Road Motor Vehicles." *SAE Standard J3016_202104*. 400 Commonwealth Drive, Warrendale, PA, United States: SAE International. https://doi.org/10.4271/J3016_202104.
Thornberry, Cameron L. 2014. "Extending the Human Controller Methodology in Systems-Theoretic Process Analysis (STPA)," 219.