Title
Automated Vehicle Safety: Heavy Duty Safety Drivers and Remote Operators — Safety Metric Foundations

Authors
Correa-Jullian, Camila, PhD
D'Agostino, Mollie C.
Mosleh, Ali, PhD
et al.

Publication Date
2026-08-01

DOI
10.7922/G2VM49N1

RESEARCH REPORT

Institute of Transportation Studies | UNIVERSITY OF CALIFORNIA

Automated Vehicle Safety: Heavy Duty Safety Drivers and Remote Operators – Safety Metric Foundations

Camila Correa-Jullian, Ph.D., Postdoctoral Scholar, The B. John Garrick Institute for the Risk Sciences, University of California, Los Angeles
Mollie Cohen D'Agostino, M.S., Executive Director, MoSAIC/ PPA5, University of California, Davis
Ali Mosleh, Ph.D., Director, The B. John Garrick Institute for the Risk Sciences, University of California, Los Angeles
Jiaqi Ma, Ph.D., Associate Professor, UCLA Mobility Lab, University of California, Los Angeles

August 2026

California Resilient and Innovative Mobility Initiative

Report No.: UC-ITS-2024-27-5L | DOI: 10.7922/G2VM49N1

Technical Report Documentation Page

1. Report No.	2. Government Accession No.	3. Recipient's Catalog No.
UC-ITS-2024-27-5L	N/A	N/A

4. Title and Subtitle Automated Vehicle Safety: Heavy Duty Safety Drivers and Remote Operators – Safety Metric Foundations	**5. Report Date** August 2026 **6. Performing Organization Code** ITS-Davis, UCLA ITS
7. Author(s) Camila Correa-Jullian, Ph.D., https://orcid.org/0000-0002-4622-4064 Mollie Cohen D'Agostino, M.S., https://orcid.org/0000-0002-3689-9471 Ali Mosleh, Ph.D., https://orcid.org/0000-0002-4628-1050 Jiaqi Ma, Ph.D., https://orcid.org/0000-0002-8184-5157	**8. Performing Organization Report No.** UCD-ITS-RR-25-89
9. Performing Organization Name and Address Institute of Transportation Studies, Davis 1605 Tilia Street Davis, CA 95616 The B. John Garrick Institute for the Risk Sciences, UCLA 404 Westwood Plaza Los Angeles, CA	**10. Work Unit No.** N/A **11. Contract or Grant No.** UC-ITS-2024-27-5L
12. Sponsoring Agency Name and Address The University of California Institute of Transportation Studies www.ucits.org	**13. Type of Report and Period Covered** Final Report (October 2023 – June 2025) **14. Sponsoring Agency Code** UC ITS

15. Supplementary Notes
DOI: 10.7922/G2VM49N1

16. Abstract
This research presents a risk-informed framework to derive metrics characterizing and tracking the safety performance of heavy-duty automated vehicles (HD-AVs) operations. A combination of traditional and novel hazard identification methods are leveraged to study potential human-system interactions in HD-AVs operations, ranging from safety drivers to remote operators.

17. Key Words	18. Distribution Statement
Heavy-Duty Vehicles, Automated Vehicles, Automated Vehicle Control, Human Factors, Human Machine Systems, Risk Analysis, Transportation Safety	No restrictions.

19. Security Classification (of this report)	20. Security Classification (of this page)	21. No. of Pages	22. Price
Unclassified	Unclassified	75	N/A

Form Dot F 1700.7 (8-72) Reproduction of completed page authorized

About the UC Institute of Transportation Studies

The University of California Institute of Transportation Studies (UC ITS) is a network of faculty, research and administrative staff, and students dedicated to advancing the state of the art in transportation engineering, planning, and policy for the people of California. Established by the Legislature in 1947, ITS has branches at UC Berkeley, UC Davis, UC Irvine, and UCLA.

The California Resilient and Innovative Mobility Initiative

The California Resilient and Innovative Mobility Initiative (RIMI) serves as a living laboratory – bringing together university experts from across the four UC ITS campuses, policymakers, public agencies, industry stakeholders, and community leaders – to inform the state transportation system's immediate COVID-19 response and recovery needs, while establishing a long-term vision and pathway for directing innovative mobility to develop sustainable and resilient transportation in California. RIMI is organized around three core research pillars: Carbon Neutral Transportation, Emerging Transportation Technology, and Public Transit and Shared Mobility. Equity and high-road jobs serve as cross-cutting themes that are integrated across the three pillars.

Acknowledgments

This study was made possible with funding received by the University of California Institute of Transportation Studies from the State of California through the Road Repair and Accountability Act of 2017 (Senate Bill 1) and/or a one-time General Fund allocation in the 2021 State Budget Act for the Resilient and Innovative Mobility Initiative. The authors would like to thank the State of California for its support of university-based research, and especially for the funding received for this project. The authors would also like to thank Anna Cosmin-Spanoche, Graduate Student Researcher at The B. John Garrick Institute for the Risk Sciences, University of California, Los Angeles who provided the foundations for the development of this report.

Automated Vehicle Safety: Heavy Duty Safety Drivers and Remote Operators – Safety Metric Foundations

Camila Correa-Jullian, Ph.D., Postdoctoral Scholar, The B. John Garrick Institute for the Risk Sciences, University of California, Los Angeles
Mollie Cohen D'Agostino, M.S., Executive Director, MoSAIC/ PPA5, University of California, Davis
Ali Mosleh, Ph.D., Director, The B. John Garrick Institute for the Risk Sciences, University of California, Los Angeles
Jiaqi Ma, Ph.D., Associate Professor, UCLA Mobility Lab, University of California, Los Angeles

August 2026

California Resilient and Innovative Mobility Initiative

Report No.: UC-ITS-2024-27-5L | DOI: 10.7922/G2VM49N1

Table of Contents

Table of Contents

List of Tables

List of Figures

Acronyms

ADAS Advanced Driver Assistance Systems

ADS Automated Driving Systems

AV Autonomous Vehicle

CoTA Concurrent Task Analysis

DDT Dynamic Driving Task

ESD Event Sequence Diagrams

FOC Fleet Operations Center

HD-AV Heavy-Duty Autonomous Vehicle

MRC Minimal Risk Condition

NHTSA National Highway Traffic Safety Administration

OEDR Object and Event Detection and Reaction

ODD Operational Design Domain

SAE Society of Automotive Engineers

TOR Take-Over Request

TOT Takeover Time

Executive Summary

Executive Summary

As research and development on automated driving systems (ADS) continues to advance, the full scope of safety challenges associated with autonomous vehicles (AVs) remains uncertain, particularly regarding human-system interaction. As more ADS-equipped systems move towards deployment in public roads, understanding their operational safety remains an urgent priority.

Significant efforts have focused on developing heavy-duty autonomous vehicles (HD-AVs) for various applications in the commercial freight industry, with the goal of improving traffic efficiency and safety. However, developers, fleet operators, and regulators must address the unique safety risks that ADS technology can introduce. Most visions of HD-AV operations aim for the systems defined by the Society of Automotive Engineers (SAE) as Level 4 ADS, which do not require human intervention under defined conditions, however other technological, legal, and socioeconomic factors may continue to require human oversight, often in the form of on-board safety drivers or remote operators. This ongoing human involvement presents important safety questions that cannot be fully addressed by measuring only the performance of the ADS itself.

This report presents a risk-informed framework to develop safety metrics that help characterize and track operational risks in HD-AV operations. These metrics enable system designers, operators, and regulators to better understand and manage safety throughout system development and deployment. Our approach builds on a structured risk assessment framework originally developed to assess operational safety in driverless passenger vehicles, with a focus on the roles of safety drivers and remote operators in middle-mile freight applications (between first-mile and last-mile delivery). The framework integrates tools to identify risks stemming from the interactions among hardware, software, human, and organizational elements, and aims to highlight gaps in existing safety standards through structured analysis and modeling.

To support this approach, a set of metrics focusing on key ADS tasks and human-system interactions was developed, using both traditional and novel hazard identification and modeling tools, such as event sequence diagrams (ESDs) and concurrent task analysis (CoTA). These metrics are grouped by categories such as control transitions, alerts, incidents, fallbacks, and human-ADS trust. They include quantitative indicators, such as the success rate of driver-initiated takeovers or the frequency of unacknowledged alerts, and qualitative measures gathered through interviews and questionnaires. The metrics were designed to be tracked through ADS logs and operational data during simulation, testing, and live operations, helping identify system weaknesses, improve human-machine interfaces, and inform design decisions. These are leading indicators that allow safety trends to be observed over time, even when no incident occurs, providing a proactive method to assess operational safety in HD-AV systems.

In summary, this report outlines a structured method for developing human-centered safety metrics that reflect how drivers and remote operators interact with ADS-equipped vehicles in HD-AV operations. These metrics can guide system design, improve operator training, and support safer deployments. As HD-AV fleets expand, these concepts will be essential for evaluating safety and ensuring that automation responsibly enhances, rather than replaces, human oversight. Research in this area can help decision-makers better understand the emerging risks and benefits of advanced driver assistance systems, highly automated ADS technologies, and HD-AV operations, enabling more informed evaluations of their impact on safety, labor, and the broader transportation system.

Contents

Introduction

Automated driving systems (ADS) and related technologies hold significant potential to transform transportation systems, freight logistics, and regulatory structures. Their widespread adoption, however, depends on a thorough understanding and mitigation of the risks they introduce. Among these, human-system interaction has emerged as a critical area, particularly as more systems reach public road deployment with varying degrees of automation.

One of the most prominent and promising applications of driving automation is in heavy-duty commercial transportation—with vehicles referred as heavy-duty autonomous vehicles (HD-AVs). In this context, the integration of advanced driver assistance systems (ADAS) and ADS is largely motivated by the need to reduce traffic collisions attributed to human error, improve operational efficiency, and extend service hours (Bhoopalam et al. 2023). The Society of Automotive Engineers (SAE) defines six levels of driving automation (Levels 0-5), based on the level of human involvement in the dynamic driving task (DDT) and the extent of the operational design domain (ODD) in which the system is certified to operate. At Levels 0-2, the human driver remains primarily responsible, with ADAS providing assistance under specific conditions and limited tasks. At Level 3, the ADS assumes the driving tasks within a defined ODD, however, the human driver must still act as a fallback-ready user. At Level 4, the ADS can perform the full DDT and fallback functions without human intervention, but only within a limited ODD. At Level 5, the ADS is capable of performing all DDTs and fallback functions independently, without restrictions to a specific ODD (SAE On-Road Automated Vehicle Standards Committee 2021).

While these levels offer a standardized framework for classifying automation, real-world HD-AV deployments currently reflect a more limited and varied application of these capabilities. For example, current designs for HD-AV operations today span various applications—such as middle-mile delivery, long-haul trucking, and port drayage—each with unique operational profiles and safety demands. Despite strong industry interest, public road deployment of HD-AVs remains limited. As of April 2025, Aurora launched the first fully driverless Class 8 trucks operating between Dallas and Houston on I-45 using Level 4 ADS. While this marks an important milestone, broader adoption will require ongoing assessment of the risks introduced by expanding the number of HD-AVs in service, increasing operational hours, and relying on emerging technologies.

As these deployments progress, it is essential to consider the broader implications for transportation safety. In the US, the National Highway Traffic Safety Administration (NHTSA) plays a key role in overseeing the safe operation of heavy-duty vehicles. In 2021, NHTSA reported early 14 million medium- and heavy-duty trucks were registered. Large trucks accounted for 9% of all vehicles involved in fatal crashes, despite making up only 5% of total registrations. Furthermore, 72% of fatalities in crashes involving large trucks were occupants of other vehicles, compared to 11% of fatalities being occupants of the truck (United States. Department of Transportation. National Highway Traffic Safety Administration. National Center for Statistics and Analysis. 2023). These statistics highlight the need for more reliable, safe, and efficient systems. Many developers and fleet operators are pursuing ADS as a path forward. However, efforts must be directed towards understanding and assessing the new risks that arise by introducing this technology, in particular regarding short-to-medium-term human-system interactions supporting and overseeing HD-AV operation.

These safety concerns, alongside growing industry efforts, underscore the need to better understand the risks introduced by human-system interaction in HD-AVs. Currently, there are multiple prospective models of HD-AV operations, with levels of driving automation ranging from 2-4 depending on the specific application. Many HD-AV operations continue to include humans in key roles such as onboard safety drivers, safety operators, or remote fleet supervisors, either in monitoring roles or by completing sections of the drive (Bhoopalam et al., 2023). These human agents interact with the system during different phases of operation, including pre-ODD driving, control transitions, and emergency responses. While many systems are designed to operate at SAE Level 4, in practice these hybrid operations may function more like Level 2 or 3 deployments depending on the scenario. As such, understanding how human and machine agents interact is essential for ensuring safe and effective system performance. To support safe deployment, one important step is the development and application of structured safety metrics that account for both system and human roles.

Human-Centric Safety Metrics in HD-AV Operations

Safety metrics provide a structured framework for analyzing overall system safety and are widely used across high-risk industries such as maritime, nuclear, and aviation. These metrics typically fall into two categories: lagging indicators, which are based on past incidents, and leading indicators, which reflect precursors to potential safety events (Reiman and Pietikäinen 2012).

In the context of automated driving technology, most existing safety metrics focus on functional safety; that is, whether the system performs according to specifications under normal conditions (International Organization for Standardization 2018). However, these metrics often overlook operational processes and the complexity of human-system interactions. While ongoing efforts seek to extend this perception of safety towards broader definitions, there are still key challenges to address (International Organization for Standardization 2022). For example, while crash data is commonly used in safety assessments, severe collisions are rare events. As a result, traditional lagging metrics, such as incident rates per mile driven or the number of injuries and fatalities, can offer limited insight, making it difficult to draw statistically meaningful conclusions from small datasets.

To address this gap, surrogate safety metrics have been developed to measure traffic conflicts and other incidents that are statistically linked to crashes but do not necessarily result in one (Bin-Nun et al., 2023). There is growing interest in developing leading metrics that capture safety-relevant behavior and events even when no incidents occur. In the context of ADS, some surrogate safety metrics are derived from human-driven traffic metrics, while others are based on data collected from simulator studies and real-world ADS testing (Automated Vehicle Safety Consortium 2021). These surrogate safety metrics—such as time-to-collision, deceleration rate to avoid the crash, and DeltaV—aim to quantify interactions with other road vehicles during "near misses" and non-failure events, i.e., events that do not directly lead to a crash, which may or may not involve a failure at a component or sub-system level, but that otherwise exhibit undesirable behavior (Wang et al. 2021). Additional metrics such as aggressive driving—a binary indicator of repeated, unsafe longitudinal or lateral maneuvers—are used to assess the safety profile of autonomous vehicles (Wishart et al. 2020). Collectively, such leading metrics provide further indications of system safety, enabling standardization and comparison of non-failure events before vehicles enter public deployment (Reiman and Pietikäinen 2012).

In relation to human-system interaction, several approaches and tools have been developed to assess human driver performance. At lower levels of driving automation, research has focused on the development of driver monitoring systems, integrated into many ADAS-equipped vehicles, to track physical and visual cues (e.g., head position, eye movements) to determine driver attention (Koesdwiady et al. 2017; Dong et al. 2011; Khan and Lee 2019). In addition to real-time monitoring, self-reported assessment methods, such as the driver behavior questionnaire (De Winter 2013; Reason et al. 1990) and the driving inattention scale (ARDES), can help to estimate the likelihood of attention-related driving errors (Castro et al. 2024; Ledesma et al. 2015; Castro et al. 2019).

A major focus in level 3 ADS human-system interaction research is on takeover requests (TORs) and developing adequate metrics to capture the human and system performance during these control transition events. Human factor studies in this area usually analyze reaction time, attention, fatigue, and takeover quality, all of which are relevant for HD-AV operations (Petermeijer et al. 2017). Other studies emphasize the increased cognitive load and task complexity associated with emergency scenarios, which reinforce the need for detailed and well-calibrated safety metrics (DeGuzman et al. 2021; Figalová et al. 2023; Payre, Perelló-March, and Birrell 2023). Several studies have explored trust calibration, conflicts of control authority, regaining and maintaining situation awareness, take-over control mechanisms, and the design of human-system interface (Chu et al. 2023; Vanderhaegen 2021; Warg, Skoglund, and Sassman 2021). These studies are generally conducted in driving simulator environments for partial (level 2) and conditional (level 3) automated driving.

Driving simulators, such as OpenCDA and CARLA, offer a valuable platform to study aspects of driving, including vehicle functions and behavior, as well as human factors. These tools are relatively low-cost, safe, and effective for simulating rare or high-risk events that are difficult to study in the real world (Dosovitskiy et al. 2017; Xu et al. 2021; 2023). In addition to capturing vehicle performance, simulators can be used to evaluate human behavior in response to ADS functions. Common metrics used to evaluate the human's performance in take-over scenarios include the reaction time to the take-over request (TOR) and the take-over quality elements, e.g., take-over time (TOT), lane-keeping performance, minimum time-to-collision, deceleration frequency, jerk, and steering smoothness.

Cognitive effects on driver take-over performance (reaction time and quality) commonly studied include workload, drowsiness, fatigue, situation urgency, and available response time (B. Zhang et al. 2019). Many of the issues addressed by these metrics are key to characterize safety drivers involved at different stages of development, testing, and operation, and also apply to highly automated (level 4) driverless vehicles whose operation is supervised, to some extent, by remote operators. Remote supervision introduces additional complexity due to potential latency in wireless communication and limited perception. The challenges these operators face differ substantially from on-board drivers, principally given the physical disconnection to the vehicle. Thus, remote operators must rely entirely on the human-system interface provided and the information transmitted by the vehicle, both of which are subject to external network latency issues (T. Zhang 2020; Kuru 2021). These aspects of remote operation increase the already difficult task of maintaining situation awareness over extended periods of time, or if lost, of quickly regaining situational awareness, particularly in emergency situations (Tener and Lanir 2022) while also monitoring multiple vehicles simultaneously (Bogg and Birrell 2025).

Despite growing attention to the impact of human-ADS interactions on overall system and operation safety, research on safety metrics that evaluate the *quality* of human-ADS interactions remains underdeveloped. To effectively assess human-ADS interactions, developers should adopt a multi-method approach, combining quantitative measures (e.g., response time, maneuver quality) with qualitative data (e.g., post-scenario interviews or surveys). TOR metrics, for example, reveal not only how quickly a driver responds but also how successfully they resume control—information that is critical to system design and operational safety (DeGuzman et al. 2021; Morales-Alvarez et al. 2020; B. Zhang et al. 2019). These gaps point to an urgent need for the development of standardized, validated metrics across diverse use cases. A combination of quantitative and qualitative methods can offer a more holistic understanding of safety. **Figure 1** summarizes common data sources for both human and ADS safety metrics.

More research is needed to define, standardize, and validate such metrics across different system architectures and use cases—including different configurations of drivers, remote operators, and levels of driving automation.

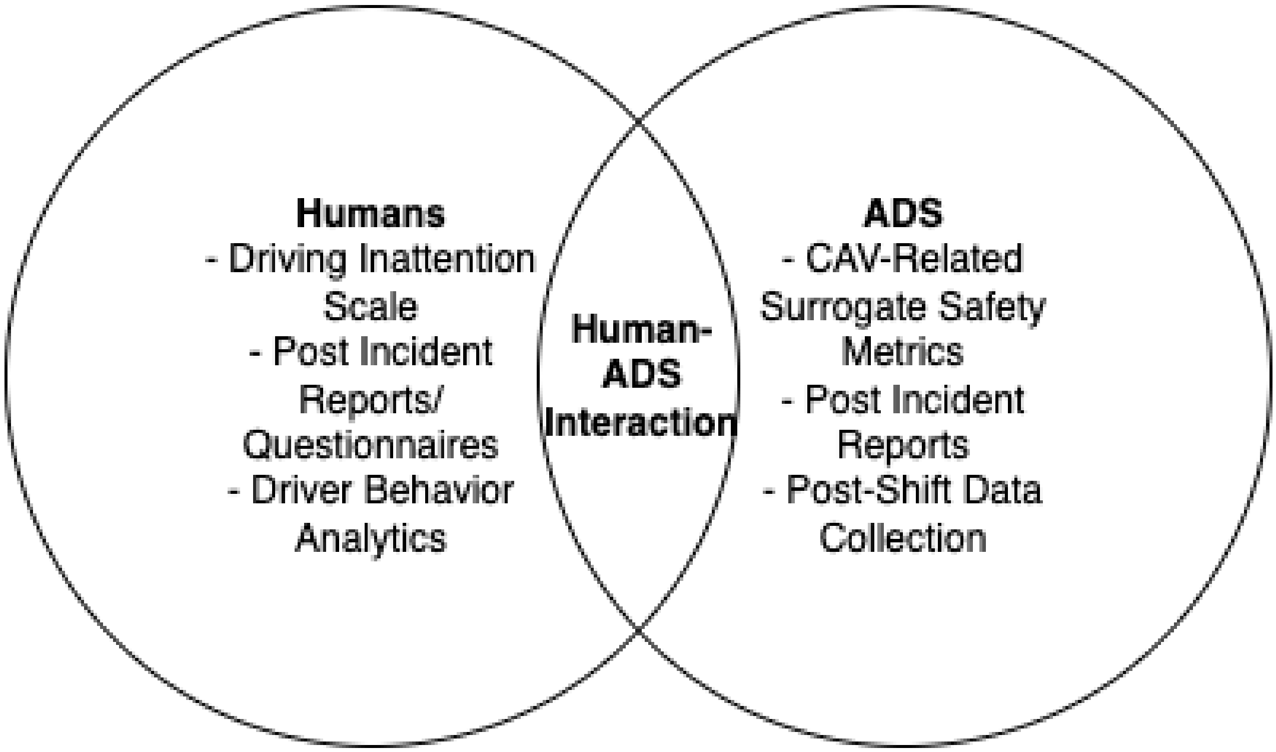


Figure 1. Depiction of Metric Sources for Humans and ADS

Report Organization

To support the safety assessment of HD-AV systems, a set of metrics that includes the unique aspects of human-system interaction must be defined. To address this need, this work focuses on assessing the impact of human-system interaction on operational safety in HD-AV deployments. A structured hazard identification framework is used to derive metrics that capture key interaction points between humans and the ADS.

This report is structured as follows. A general reference fleet, representing most commercial HD-AV operations, is developed to establish a system breakdown and define operational phases based on system interactions. The system is then further decomposed using scenario-based modeling and critical task identification, applying a combination of traditional and novel hazard identification and risk modeling tools. The developed models focus on the most critical interactions between humans and the ADS, using this information to construct relevant metrics. These metrics are essential for evaluating not only the functional performance of the ADS but also the safety and reliability of the broader socio-technical system in which it operates.

Safety Metric Derivation Framework

This work follows a scenario-based approach to conduct a safety risk analysis of human-system interactions in the context of HD-AV operations. High level operational scenarios are selected for an in-depth risk analysis, guided by the typical dimensions of risk:

i) what can go wrong (development of risk scenarios);

ii) how likely it is that that will happen (expressed qualitatively or quantitatively);

iii) what are the consequences in case it happens.

An operational safety hazard identification method, originally developed for complex socio-technical systems, is employed to develop new metrics for human-system interaction. This structured process combines several risk assessment techniques to analyze the interactions between agents in a complex system and identify hazards. It was initially developed to study human-system interactions in ADS operations, specifically for the case of level 4 fleets used in mobility-as-a-service (MaaS) passenger transport providers.

The hazard identification method is divided into three stages: system modelling, scenario modelling, and hazard identification (Correa-Jullian et al. 2024c). The present work uses the first stages of the hazard identification method for the HD-AV scenario, applying the event sequence diagram (ESD) and concurrent task analysis (CoTA) steps to derive metrics specifically for human-ADS interactions:

- Stage 1 – System Modeling:
  - Step 1: Define agents and their high-level tasks.
  - Step 2: Define operational phases and the transitions between them.
- Stage 2 – Scenario Modeling:
  - Step 3: Document operational phases through ESDs
  - Step 4: Model tasks and interactions using CoTA.

Stage 1 was implemented to develop a preliminary set of human-interaction metrics based on the functional breakdown and operational phases of the HD-AV system (Cosmin-Spanoche, Correa-Jullian, et al. 2024). In this work, Stage 2 is carried out to further identify and refine the metrics using a structured, task-oriented approach (Cosmin-Spanoche, Correa-Jullian, et al. 2024). Further details on the method and extended results can be found in (Correa-Jullian et al. 2024c). A depiction of the initial tasks used to derive metrics and hazard scenarios is shown in **Figure 2**.

The first two steps of the method involve modelling the system by performing a breakdown for each agent and defining operational phases. A reference profile of HD-AV operations and roadway conditions was defined based on literature review and stakeholder engagement, following NHTSA guidance for ADS system design.

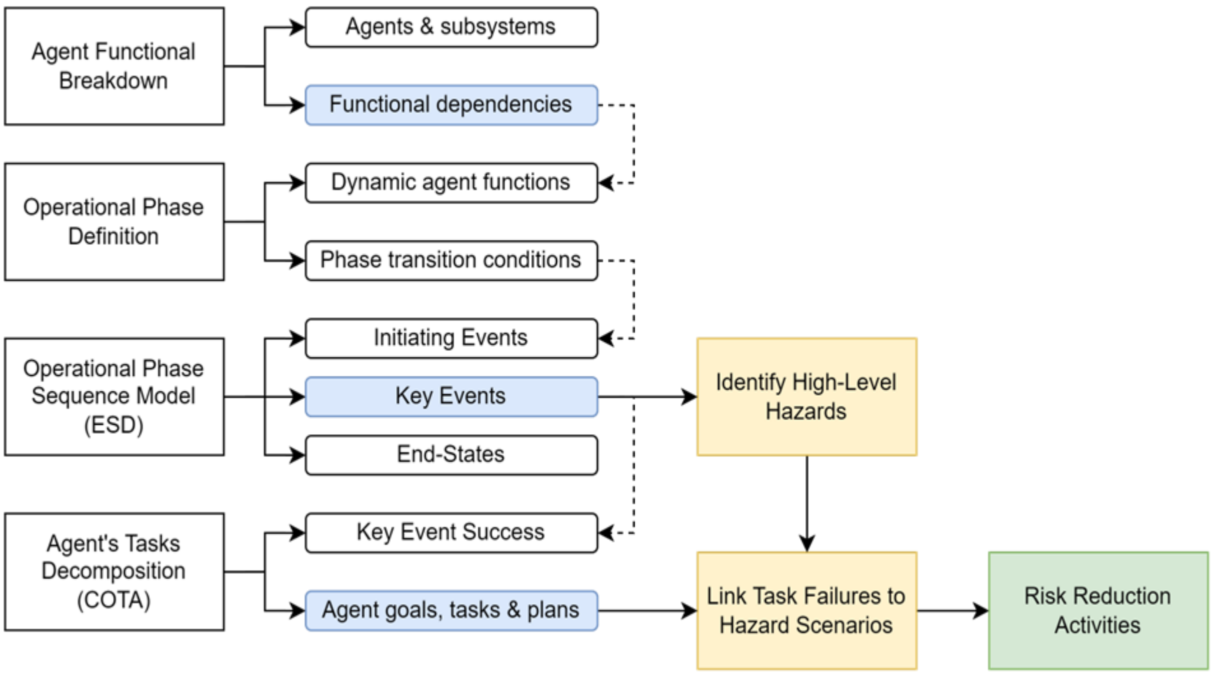


Figure 2. Operational Safety Hazard Identification Framework

The HD-AV systems include hardware, software, and human agents, according to the operational scenario defined. The human operators considered are:

- Onboard safety drivers, who may share the driving tasks with the ADS in certain conditions (e.g., in the event the system leaves the operational design domain or experiences a scenario where a clear trajectory cannot be computed).
- Remote operators and assistants, who can monitor the operation and, when necessary, support the operation, such as providing waypoints or intervening to take control of AVs.

The reference system is then analyzed from the perspective of the human-system interactions, focusing on different human-ADS configurations. The goal of this analysis is to determine safety-relevant human-system interactions and propose safety metrics to assess and track them. The developed models focus on the most critical interactions between humans and ADS, using this information to construct relevant metrics, and informing the design of experiments conducted in a driving simulation environment.

Hazard Identification Methods

Concurrent task analysis (CoTA) was employed to investigate the role of human-systems interactions on the system safety. This is a modern hazard identification method developed to study emerging complex system interactions from a human-system interaction perspective. This approach allows researchers to identify aspects related to operational safety, expanding from the traditional functional safety perspective. This includes tasks associated with the CAV and humans in each scenario, as well as the functional dependencies between their tasks. Based on an initial functional model of human-automated agent interactions in AV operations, CoTA was used to model and analyze the scenarios where humans can avoid potential dangers when intervention is triggered and identify the key tasks involved and conditions for success.

CoTA was originally introduced in the context of the human-system interaction in autonomy framework, initially applied to analyze autonomous maritime vessels with remote operators (Ramos et al. 2020a; 2020b). This method can be used to analyze a system's expected behavior and performance based on decomposing system-level goals into sub-goals. These sub-goals are hierarchically organized through plans, indicating the order in which these must be performed to achieve the system-level goals, depicted in **Figure 3**. The redescription of goals into subgoals follows an extension of the cognitive model IDA (Information, Decision, and Action) to human and automated systems (Chang and Mosleh 2007; Ramos et al. 2020a; 2020b).

CoTA models were constructed from scenario analysis and event sequence diagrams (ESDs) (depicted in **Figure 4**), where key events were assigned to agents and translated into tasks and subtasks. The strength of this approach arises from the fact that hardware, software, and human elements must perform certain tasks in coordination to achieve system-level goals. Hence, it is a potent method to identify system dependencies and trace failures through sequential, parallel, or trigger tasks of multiple agents.

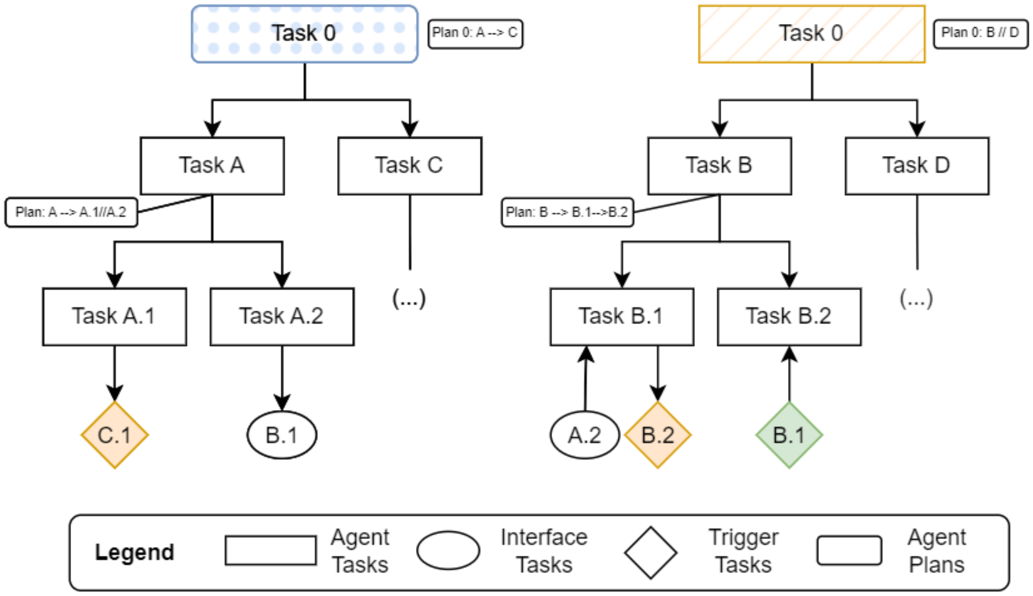


Figure 3. Example of a Concurrent Task Analysis Diagram

Hazard identification stemming from CoTA focuses on human failures and human-system interaction failures, and it provides insight into the factors that affect human performance. Hence, this can be employed to derive a list of potential failure modes and the contributing factors present in the identified hazard scenarios (Correa-Jullian et al. 2023). Key operational functions or tasks aimed at preventing or mitigation of these hazard scenarios will be identified for critical safety risks. The likelihood of occurrence and the severity of consequences of the identified hazards (i.e., task failures and unsafe control actions) was addressed qualitatively, leveraging expert knowledge and published research and standards of similar systems.

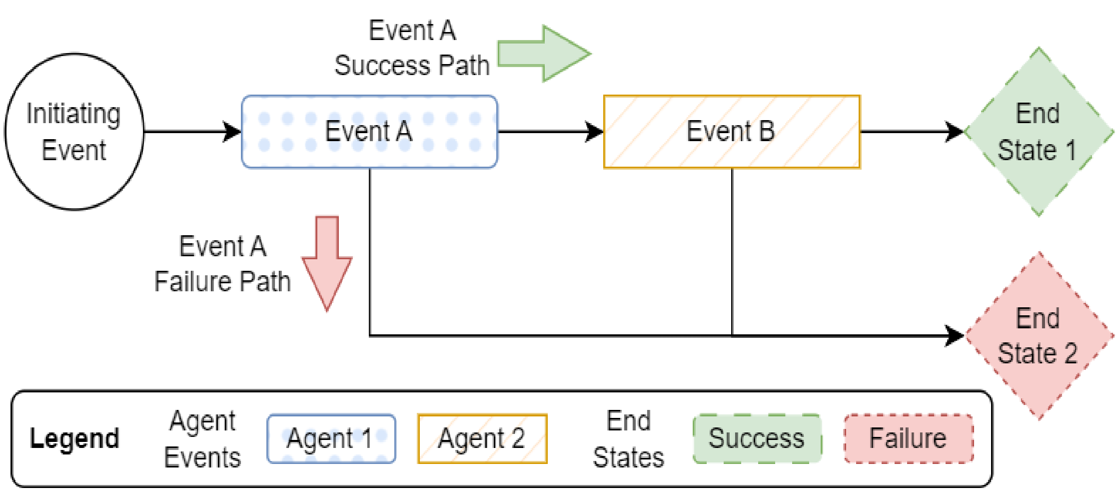


Figure 4. Example of an Event Sequence Diagram

Safety Metric Construction

While current discussions about safety drivers in HD-AVs are limited to testing phases before public deployment, safety drivers are likely to remain present during nominal (i.e., non-emergency) operations due to legal and regulatory requirements. Guidelines for creating and implementing ADS safety and performance metrics are detailed in standards such as UL4600, ISO26262, and SOTIF (International Organization for Standardization 2018; 2022) (Underwriters Laboratories 2023). However, these do not fully consider the role of safety drivers in the current planned use cases of HD-AVs. Therefore, a systematic model-based approach is needed to construct metrics related to human-system interactions in HD-AV operations.

To support the identification and mitigation of safety risks in highly automated driving systems, this work leverages hazard identification methods to derive risk-informed safety metrics. This approach builds on the analysis of failure paths derived from the ESDs and the identification of task-level failure modes using the CoTA. The process began by examining potential failure scenarios, particularly those involving human failures and human-system interaction breakdowns, as identified in the developed ESDs. These scenarios provided insight into the contributing factors that influence human performance and system reliability. From these, potential failure modes and unsafe interactions were systematically extracted based on the CoTA structure. Key operational functions and mitigation tasks associated with these hazard scenarios were then identified to determine which aspects of human-system interaction are most critical to operational safety. Qualitative assessments of hazard likelihood and severity were incorporated based on expert judgment, published literature, and applicable standards for similar systems. This allowed for prioritization of risks and identification of areas where metrics could be most impactful for safety assurance.

To quantify and monitor these critical aspects, a set of interaction metrics was developed. Metrics were defined by assessing which measurable elements of the identified tasks could indicate vulnerabilities or latent conditions in the system. The resulting metrics reflect both the contributing factors to unsafe scenarios and the effectiveness of key safety-related tasks. The metrics were developed based on observations of how specific tasks contribute to failure scenarios and were organized into categories reflecting different aspects of human-system interaction. By examining the contributing factors present in the identified hazard scenarios, and by identifying critical tasks aimed at preventing or mitigating these scenarios, the derived metrics offer a targeted means of evaluating operational safety.

Deriving Safety Metrics for Human-System Interactions in HD-AV Operations

Reference HD-AV Fleet Operations

The reference fleet was defined to represent the anticipated operational configuration of short-term to medium-term heavy-duty automated vehicle (HD-AV) fleet operations. Extensive research was conducted by examining relevant publications from authoritative sources such as NHTSA, AVSC, and SAE International (Thorn, Kimmel, and Chaka 2018; Blanco et al. 2020; Chaka et al. 2021; Automated Vehicle Safety Consortium 2021; 2023; 2019). This review was supplemented with voluntary safe self-assessments published by various manufacturers and developers involved in HD-AV operations based on the outline provided in NHTSA's 2017 Voluntary Guidance (National Highway Traffic Safety Administration 2017).

The reference fleet's definition revolved around its operational profile, characteristics, and known issues of the human and ADS agents involved in HD-AV operations. Specifically, the literature review focused on HD-AV projects at various stages of development to determine:

- Operational profiles: roadway conditions and operations, i.e., middle-mile operations, drayage operations, urban delivery; geographical, weather, road types and conditions included in ODDs; usage profile, planned routes, connectivity requirements.
- System functionality: safety driver functions, vehicle characteristics and HD-specific technology, AV functions and technology, i.e., level of automation, remote support functions, safety-related functions. Emphasis was placed on detailing the interactions of the HD-AVs with safety drivers and remote fleet operators.

Traditional Heavy-Duty Freight Operations

Heavy-duty vehicles are defined as Class 7-8 vehicles, or vehicles weighing over 26,001 pounds (11,794 kg). The classes are denoted by the limitations of Class 7 being between 26,001 and 33,000 pounds (11,794 to 14,969 kg), and Class 8 being over 33,000 pounds (14,969 kg). Heavy-duty vehicles serve various roles: for example, as construction vehicles (such as dump trucks and cement mixers), garbage trucks, and large transit buses. However, their most common use is in commercial transport operations. Predominantly, heavy-duty vehicles are used in commercial trucking operations, which can be either short-haul or long-haul, and account for approximately 70% of the annual freight tonnage in the United States (Talebian and Mishra 2022). These are further categorized into drayage, freight, less-than-truckload, and intermodal trucking, among other distinctions.

Commercial trucking corporations are primarily structured under two different paradigms. The first model is led by owner-operators, who are independent contractors that lease their services to trucking companies or directly to shippers. The second consists of fleet operators, which own and operate their own fleet of trucks and are responsible for hiring drivers to operate these vehicles. Operators of a commercial truck must possess a valid commercial driver's license and must undergo extensive background checks and training prior to operation. The Federal Motor Carrier Safety Administration regulates standards relating to commercial trucking, including creating appropriate commercial driver's license tests and enforcing regulations.

Various reports have evaluated the impact of replacing human drivers with automated technologies, estimating crash reduction rates from 50% (Shetty et al. 2022) to 90% (Bonnefon 2021). However, this estimate does not imply that the drivers are at fault in all the associated collision scenarios nor that removing the driver will entirely decrease crash rates (Zhai et al. 2023). Furthermore, the operational hours of commercial trucking are currently restricted by the Hours-of-Service regulation of the U.S. Department of Transportation (USDOT). Increased automation could potentially extend these hours towards continuous 24/7 operation. Additionally, incorporating platooning and live traffic data could enhance fuel efficiency, thereby reducing operational costs (Talebian and Mishra 2022).

Currently, there are multiple prospective models of HD-AV operations, with levels of driving automation ranging from 2 to 4 depending on the specific application. Current companies involved in this industry at different stages of development include Aurora, Kodiak Robotics, Torc Robotics, TuSimple (which ceased operations in the United States in late 2023), and Ike (acquired by Nuro). Many operational models contemplate interactions with human operators and safety drivers, either in monitoring roles or by completing sections of the drive. For instance, shorter-term applications such as middle-mile operations consider safety drivers to complete sections of the driving operation prior to entering and after exiting the Operational Design Domain (ODD).

In addition to the safety driver, these commercial operations may also consider the active participation of a fleet operations center, and, in some cases, an additional onboard safety operator. The human operators are not only expected to interact with the HD-AV during planned sections of the operation, but also to collaborate with the vehicle in emergency situations, manage planned and unplanned control transitions, and conduct post-incident procedures. These vehicles operate on highways and on/off ramps, which are denoted the "middle-mile" for commercial goods transport, as depicted in **Figure 5**. In contrast to first-mile and last-mile operations that require human drivers to complete certain tasks, middle-mile operations lend themselves better for automated driving. However, this duality implies that HD-AVs are expected to operate in manual driving and automated driving mode on a nominal basis.

HD-AV Reference Fleet

For this work, a reference fleet was defined based on a current sample of companies developing and testing in HD-AVs. Characteristics such as ODD restrictions, vehicle sizing, and human-system interaction types were selected to construct a representative model of HD-AVs in the industry. In addition, guidance from NHTSA for the design of ADS systems was employed. Although many existing HD-AV companies have significant differences in their designs and processes, this general case is representative of the designs and processes of the short-term development of this industry.

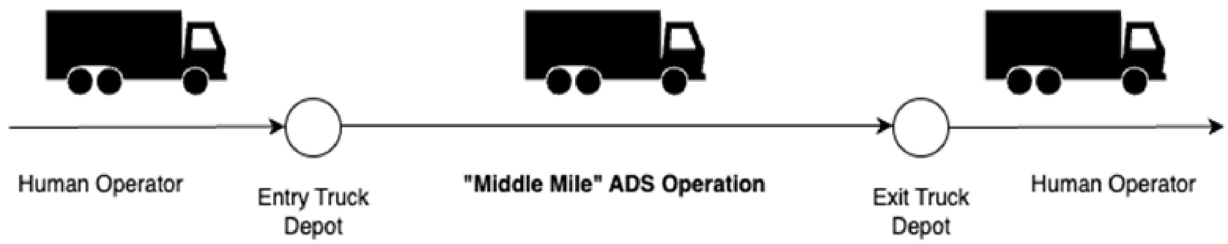


Figure 5. Depiction of ADS Middle-Mile Operations

Based on the literature review and the team's previous research experience, the analysis focused on operational scenarios involved in middle-mile highway operations that employ retrofitted Class 8 commercial vehicles owned by fleet operators and procured from an external ADS developer or vehicle manufacturer. Therefore, the primary responsibility of the fleet operator is to ensure the proper and safe functioning of the fleet, following the technical requirements set by the ADS developer and additional HD-AV guidelines. The fleet operator may be also responsible for training safety drivers and/or remote operators, as well as monitoring operations through a fleet operations center (FOC). The reference fleet's characteristics are summarized in **Table 1**.

The reference HD-AV fleet operates in a restricted ODD with location constrained by geofenced maps. The ODD is also restricted by weather conditions, with operations taking place in high visibility and fair-weather conditions, with the most severe weather condition being light rain. While the concept of dedicated autonomous lanes has been proposed, HD-AVs are expected to operate in mixed traffic scenarios.

Table 1. Characteristics of Proposed Reference Fleet

Operational Profile	
Usage	Commercial goods transport.
Vehicle Ownership	Vehicles are owned by the fleet operators, procured from ADS developers.
ODD Restrictions	Highway “middle-mile” operations; fair weather; high visibility conditions; clear to mild rain.
Remote Fleet Operations	Backup support role, accident management, communication/warnings with safety driver and safety operator.
Fleet Operator Role	Responsible for operation and maintenance of HD-AV fleet, training safety drivers and/or remote operators (in coordination with the ADS developer, as required).
Vehicle Configuration and Capabilities	
Vehicle Segment	Retrofitted Class 8 commercial vehicles (>33,000 lbs./14,969 kg).
ADS Capabilities	Nominal Level 4 ADS. Perform all required DDTs (perception, planning, execution), including fallback to minimal risk conditions. Notify safety drivers and/or remote operators when approaching ODD limits.
Human-System Interaction Models	Trained onboard safety drivers perform driving outside the ODD and may intervene in the DDT fallbacks. Remote operators may provide additional driving assistance functions and post-incident management activities.
Inspection and Maintenance	Pre-shift inspection checklists and regular maintenance activities, with added maintenance and inspection on new sensors introduced for ADS.

In terms of driving automation capabilities, many developers envision vehicles equipped with level 4 ADS capabilities, such that safety drivers only complete the driving tasks beyond the “middle mile,” and remote driving assistance provides additional support for navigation changes, oversees fallback scenarios, and post-incident management. Under normal operating conditions, the level 4 ADS vehicle is responsible for initiating fallbacks autonomously if triggered by ODD exits, vehicle safety-critical failures, and other emergency situations. In the event the HD-AV exits the ODD under unplanned circumstances, a DDT fallback is triggered until a minimal risk condition is achieved (SAE On-Road Automated Vehicle Standards Committee 2021). In addition, operations consider the presence of a trained safety driver onboard and, in some cases, a trained safety operator, to oversee operations and intervene in the vehicle’s operation for safety reasons (Aurora 2021; TuSimple 2021; Kodiak.AI 2020; Torc 2021).

Stage 1: System Modeling

This task develops a functional model of the interactions between the human and automated agents involved in AV operations. From the reference HD-AV profiles, characteristics, and operational profiles defined in the previous task, a functional breakdown of the human, software, and hardware subsystems is performed, depending on the operational profile and roadway conditions selected.

Step 1: System Breakdown

To analyze the reference fleet's operations, the operations of the HD-AV system are broken down into different functional agents. The HD-AV system includes four main agents: the ADS, safety driver, safety operator (optional), and fleet safety operator. These agents and their main task categories are defined in **Figure 6** and described in depth in the following sections.

ADS-equipped Heavy-Duty Vehicle

The first agent present in the HD-AV system is the ADS itself, which is the software and hardware responsible for performing the DDT within the limits of the ODD. The ADS has a nominal Level 4 of driving automation, which means that it is designed to function without the need for a human to take over the vehicle while operating within the ODD. The ADS can perform the DDT fallback to reach minimal risk condition (MRC in **Figure 6**) if required (SAE On-Road Automated Vehicle Standards Committee 2021). However, current operations still require a safety driver to perform sections of transit outside of the ODD (i.e. before and after middle mile operations) and provide backup in emergency cases. In addition, the ADS contains a driver monitoring system to assess driver attention (Horberry et al. 2022; Ayas et al. 2023). A description of the ADS high-level tasks is shown in **Table 2**.

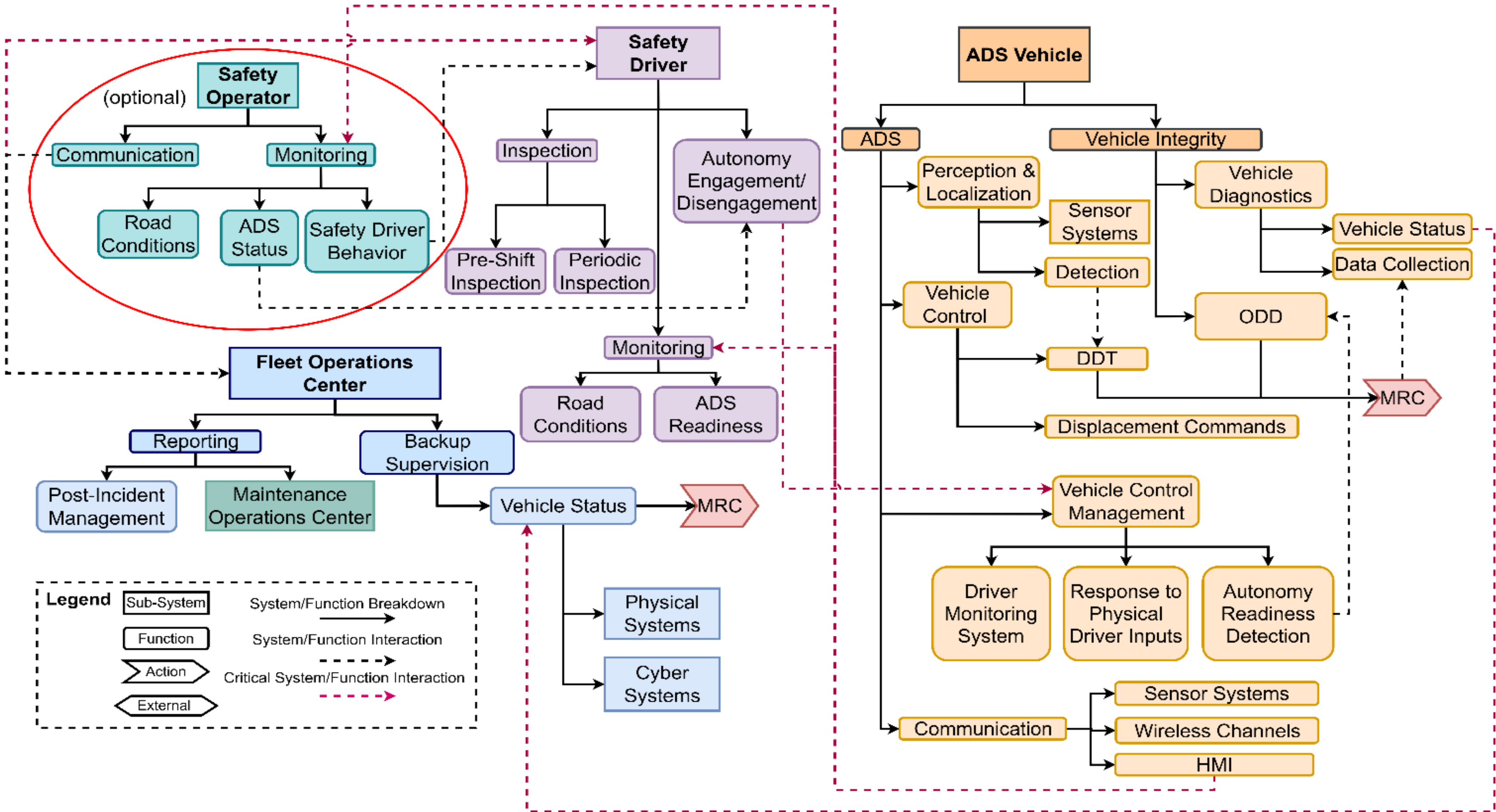


Figure 6. System Functional Breakdown

Table 2. ADS Task Categories and Descriptions

Task Category	Description
Inspection	Collect and process sensor input data coming from GPS, camera, radar, and LiDAR systems.
Vehicle Control	• Plan and implement the DDT (Dynamic Driving Task) while the vehicle is under computer control. • Issue actuation commands to the vehicle, including steering, throttle, braking, and indicator commands. • Respond to physical inputs from Safety Drivers. • Assume the fallback state when vehicle begins to exit ODD.
Vehicle Control Management	• Determine autonomy readiness based on road conditions and ODD. • Monitor Safety Driver behavior through driver monitoring system. • Inform Safety Driver/Operator about automated system state.
Communication	Communicate with the Safety Driver and Operations Center if there is an unexpected event while ADS is activated.
Vehicle Diagnostics	Assess and report status of vehicle subsystems, both related to ADS and non-automated systems.

On-board Safety Driver

Under current assumptions, safety drivers will likely be a commercial vehicle operator who possesses a valid commercial driver's license and has undergone training for commercial driving operations and interactions with the built-in ADS. For instance, it is expected that the safety driver will be highly-trained in identifying the ODD requirements, ADS limitations, and is instructed on emergency procedures (Bhoopalam et al. 2023). Their high-level responsibilities include driving the vehicle outside its ODD and engaging and disengaging the Automated Driving phase. The safety driver's high-level tasks are described in **Table 3**.

Table 3. Safety Driver Task Categories and Descriptions

Task Category	Description
Inspection	• Conduct pre-trip inspection of safety-critical vehicle systems. • Inspect the vehicle and trailer every time the truck stops.
Monitoring	Monitor road and vehicle behavior.
Autonomy Control Transitions (Engagement/ Disengagement)	• Engage and disengage the vehicle's autonomy system. • Take control of the vehicle in case of a disengagement. • Manually drive the vehicle when it is outside of its ODD.
Communication	Communicate with the Operations Center about issues.

Control transitions between the human and the ADS occur when there is a limitation of the ADS capabilities at certain levels of autonomy. A control transition can allow the driver to take control of the vehicle and operate it outside of its ODD, considering, for example, geographic or weather restrictions, vehicle or sensor failures, and encounters of traffic scenarios outside of the ODD. Additionally, while the vehicle is within the ODD, a control transition can allow the human to transfer control of the driving task to the ADS.

As a convention for this work, the transferring of control from the driver to the ADS is referred to as a "handover"; the reverse transfer from the ADS to the driver is referred to as a "takeover." The driver initiates a control transition through physical mechanisms, either by directly engaging with the steering wheel, brakes, or accelerator pedals or by requesting a handover through buttons installed on the steering wheel. For a handover or takeover event to occur, the HD-AV system design must clearly include the order and coordination of tasks needed to transfer control.

Safety Operator

The safety operator is an additional human agent onboard the vehicle in the passenger's seat whose responsibility is to monitor road conditions and the state of the HD-AV. These operators may interact with dedicated human-system interface display to identify any potential issues internally with the ADS or externally to warn the safety driver. In addition, they serve as a party that enables communication between the safety driver and fleet operations center. In the event the fleet operations do not include a safety operator, the safety operator's tasks are incorporated into the safety driver's responsibilities. A description of the safety operator's high-level tasks is shown in **Table 4**.

Table 4. Safety Operator Task Categories and Descriptions

Task Category	Description
Communication	• Communicate ADS intentions, status, and misbehavior to safety drivers. • Communicate with the operations center.
Monitoring	• Monitor the operation of the ADS via human-system interface internal display. • Monitor human-system interface for missed detections, false detections, unsuitable motion plans, and poor data quality. • Warn safety drivers to disengage ADS. • Record notes about system and road conditions and incidents for post-shift debrief.

Fleet Operations Center (FOC)

The fleet operations center is a physical space in which hired operators monitor the HD-AV fleet in a control room environment. Each fleet operator may be tasked to monitor multiple HD-AV systems through a dashboard and provide warnings to safety drivers and/or safety operators on board. In addition, fleet operators receive incident notifications automatically and play a role in route planning and accident management. A description of the high-level tasks of the operators at the fleet operations center is shown in **Table 5**.

Table 5. Fleet Operations Center Task Categories and Descriptions

Task Category	Description
Backup Supervision	• Perform a backup support role for safety driver and safety operator. • Monitor fleet of vehicles and their statuses from control room environment via dashboard.
Reporting	• Communicate with safety driver/safety operator about potential obstacles and risks coming ahead. • Respond to alerts about safety driver inattention. • Warn safety drivers to disengage ADS. • Respond to accident scenarios, alert emergency services, and participate in post-accident debrief.

A summary of the overall HD-AV high-level tasks and their divisions are compared in **Table 6** and **Table 7**. As shown, the ADS and the driver divide the DDT and related tasks and coordinate this task allocation based on control transitions (Correa-Jullian et al. 2024b). The high-level tasks are described in **Table 6**. Four levels of engagement are used to categorize the participation of the agents in each of the high-level tasks:

- Always: This task is continuously performed or available during the system's operation
- Partial: This task is performed temporarily (while the ADS or Manual Driving is engaged) or up to a partial degree (the remote operator is engaged with multiple vehicles)
- Backup: This is a safety-backup task only performed if another agent has failed to perform a main or temporary task
- Never: The agent does not perform this task during operation.

Table 6. HD-AV High-Level Task Descriptions

High-Level Task	Description
Monitor DDTs	Perform object and event detection tasks.
Plan DDTs	Perform the planning stage of the object and event reaction tasks.
Execute DDTs	Perform the execution stage of the object and event reaction.
Control Vehicle	Physically control the vehicle.
Monitor Teammate	Monitor the state of the driver/operator.
Request Support	Request support from other agents to perform shared tasks.
Initiate Control Transitions	Initiate a handover or takeover request.

Table 7. HD-AV Task Division Summary

Tasks/Agent	ADS	Driver	FOC
Monitor DDTs	Always	Partial	Backup
Plan DDTs	Partial	Partial	Never
Execute DDTs	Partial	Partial	Never
Control Vehicle	Partial	Partial	Never
Monitor Teammate	Always	Always	Partial
Request Support	Partial	Backup	Backup
Initiate Control Transitions	Always	Always	Never

Step 2: Operational Profile

The operation of the reference fleet and the functions each of the human and machine agents perform may vary depending on certain conditions. For this work, the high-level tasks of the agents described in the previous section are organized into an operational phase diagram depicted in **Figure 7**. A brief description of the five phases is presented specifically for middle mile HD-AV operations. The operational profile and further modeling stages do not include the safety operator as these are not present in most planned commercial HD-AV operations.

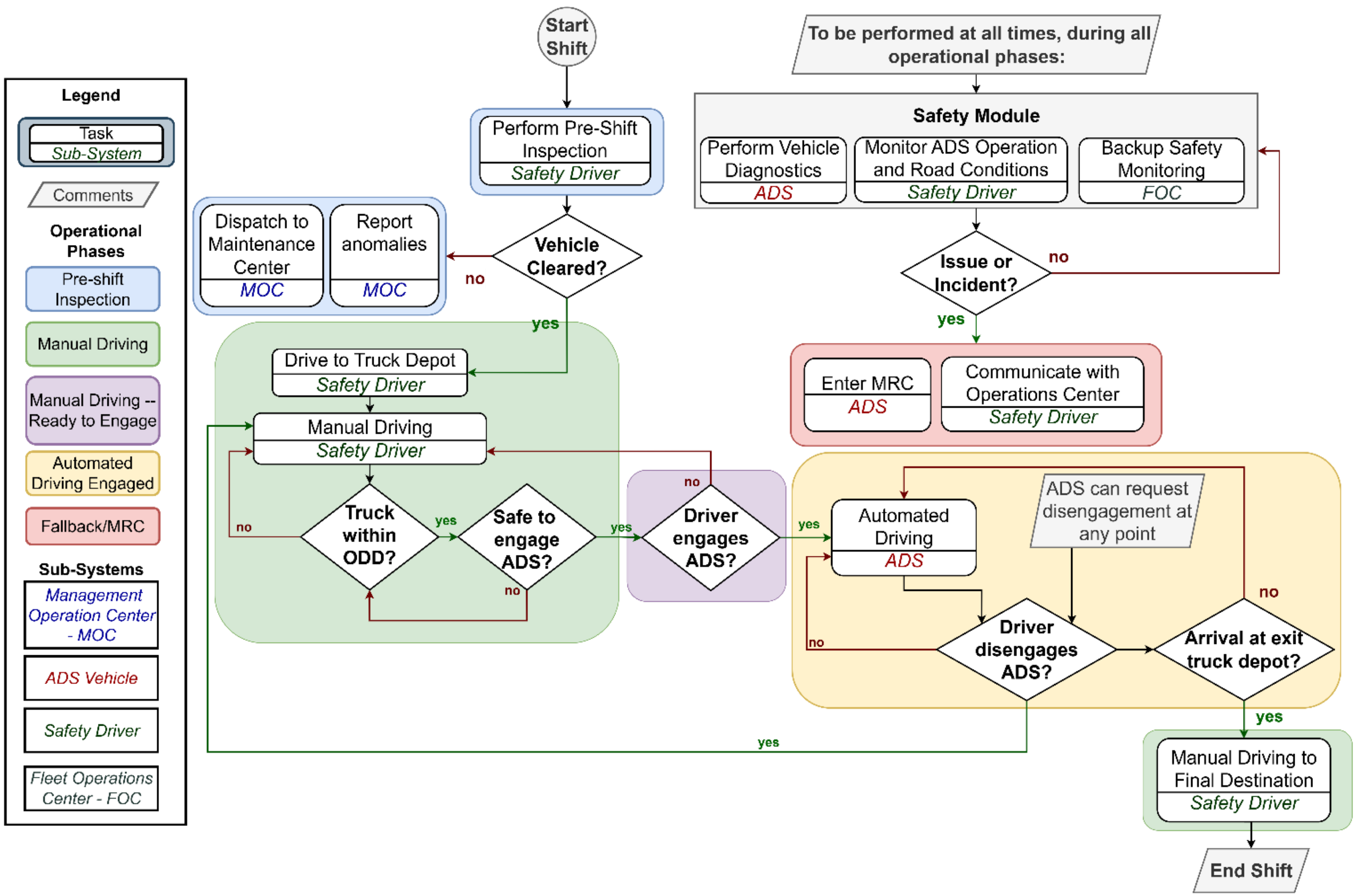


Figure 7. Diagram of Operational Phases and Subsystems

Pre-Shift Inspection

The first operational stage of a HD-AV is a pre-shift inspection, which is conducted by the safety driver prior to beginning the driving shift. This pre-shift inspection is typical for commercial trucks and involves inspecting safety-critical components (i.e. brakes, fluid levels, and tire pressure) and tracking this analysis through a standardized checklist. With the introduction of the ADS, this pre-shift inspection also includes a thorough inspection of the retrofitted ADS sensors, connectors, and mounts. Additionally, the safety driver is required to conduct a similar inspection every time the truck stops, for instance: during refueling, mealtimes, and breaks. If the vehicle is not in adequate condition, it is reported to the fleet operations center and the vehicle is dispatched to the maintenance operation center (MOC), where personnel must conduct any tests and/or repairs that will render the vehicle operational. If the vehicle is in adequate condition, it is then approved to move on to begin the shift and proceed to manual driving.

Manual Driving

The first section of the driving stage is Manual Driving. The Manual Driving phase refers to the stage where the safety driver takes full control of the DDT while it is outside the ODD. This stage can involve navigating non-highway roads leading to the truck depot, or any stage during the middle-mile journey in which the truck is not in its ODD. The fleet operations center monitors the vehicle's location, status, and any upcoming obstacles on the road. This stage can be interrupted by any incidents outside expected operations, such as system failures or unplanned route changes. In this stage, the safety driver is responsible for fallbacks and incident management. If an incident occurs, this stage is interrupted, and the fleet operations center is notified. The driver monitoring system subcomponent of the ADS monitors the drivers to ensure that they are fully aware and attentive to the driving task at hand.

Manual Driving—Ready to Engage

Manual Driving—Ready to Engage denotes when the vehicle has entered its ODD and the ADS has determined that it is feasible to activate the Automated Driving phase, but the safety driver remains in control. This stage is communicated to the safety driver through a combination of visual and auditory messages. In this stage, the safety driver decides whether to engage the Automated Driving. This stage can also occur if the safety driver disengages Automated Driving for any reason while the vehicle remains in its ODD. The driver has the option to reengage it while still in the Manual Driving—Ready to Engage phase. The fleet operations center continues to monitor the vehicle's location, status, and upcoming obstacles. As with the Manual Driving phase, the safety driver is responsible for fallbacks and incident management. If an incident occurs, this stage is interrupted, and the fleet operations center is notified.

Automated Driving Engaged

This stage occurs when the ADS is performing the dynamic driving task and is controlling the steering, brake actuation, and road monitoring. The vehicle enters this stage when the safety driver activates the

Automated Driving function via a button built into the steering wheel. In this stage, the safety driver remains in the car and monitors road conditions for any unexpected scenarios. While the vehicle is in this stage, the ADS is responsible for the fallback, but the safety driver can also intervene and take manual control at this stage. For instance, this can occur when the triggers for fallback are not autonomously detected. The ADS at this stage also detects when the vehicle is nearing the limits of its ODD and alerts the driver with visual and auditory cues in this case, notifying them to resume manual control. Additionally, if the ADS detects that the driver is inattentive, the driver and fleet operations center are notified, and they can proceed accordingly. If the vehicle exits its ODD and the safety driver has not taken control, or if an incident occurs, then the vehicle enters the Fallback/minimal risk condition stage of operations.

Fallback/Minimal Risk Condition

The fallback stage can be triggered by several incidents, for instance: internal system failures, breaches of the ODD environment, rapid changes in weather or road conditions, or incidents with other vehicles or pedestrians on the road. In the event a DDT fallback is triggered, it is expected that the ADS plans and implements a DDT fallback strategy, and achieves a minimal risk condition—i.e., the vehicle implements a safe stop, unless the safety driver intervenes to resume Manual Driving. After a minimal risk condition is achieved, the post-incident procedures are triggered and the safety driver and fleet operator decide the course of action to take, whether that be remotely assisting the vehicle or recovering it for return to the maintenance operations center. In addition, a post-shift debrief is conducted to collect information about the fallback trigger and the result of the minimal risk condition.

Stage 2: Scenario Modeling

This task is built upon the system models defined in Stage 1 and analyzes different safety-relevant scenarios and contributing elements to task success and failures. Applying a combination of top-down and bottom-up hazard identification and modeling methodologies, such as CoTA, we examined task allocation, functional dependencies, and task hierarchy between humans and ADS agents in the AVs' environments, and we selected operational scenarios, i.e., roadway conditions and human-system interaction failures. Specific emphasis is placed on hazard scenarios based on the specific safety-related functions the safety drivers and remote operators are expected to perform for each operational profile.

Step 3: Operational Phase Models

The third step in the hazard identification framework is developing ESDs representing each operational phase and the transitions. Here the primary focus is developing qualitative ESDs representing the transitions between the Manual Driving, Ready to Engage, and Automated Driving operational phases.

The critical scenarios modelled were the safety driver requesting an ADS handover (ESD 1.1, See Appendix, **Figure 15**), the safety driver performing a take-over (ESD 1.2, See Appendix, **Figure 16**), the Manual Driving phase transitioning to the ready to engage phase (ESD 2.1, **Figure 8**), and the ADS vehicle approaching the ODD limits (ESD 2.2, **Figure 9**). ESDs 1.1 and 1.2, representing the control transition events, were adapted from (Correa-Jullian et al. 2024c) and details are provided in the Appendix. The defined end states in the ESDs, shown in **Table 8**, denote a successful trip (ES1), a delayed trip (EF2), or a collision risk (EF3). If an end state serves as an initiating event for a different ESD, these are noted as transition states. ESDs 2.1–2.2 contain transition states to ESDs 1.1–1.2 depending on whether a takeover or handover is requested. Enhanced with real-world and simulation data, quantitative probabilities and risk levels can be attached to the events and end states in the ESD.

Table 8. End States for ESDs

End State	Name	Description
ES1	Trip Completed Successfully	The HD-AV safely arrives at the destination driven by the ADS agent or on-board driver.
EF2	Trip is Delayed	The trip was interrupted by a failed takeover or handover event, leading to the driver implementing minimal risk condition or the HD-AV implementing failure mitigation strategy. This can lead to an interruption or delay in the shift.
EF3	Collision Risk	The ADS agent or the safety driver fails to correctly perform a DDT fallback, and the HD-AV is at risk of collision. This can refer to a collision with property, another vehicle, or a pedestrian.

ESD 2.1, pictured in **Figure 8** with events described in **Table 9**, initiates with the ADS vehicle being cleared to begin its shift. The safety driver is expected to correctly perform the Manual Driving task prior to the vehicle entering the ODD. If the vehicle enters the prescribed ODD, the ADS detects this entrance and whether it is safe to engage the ADS.

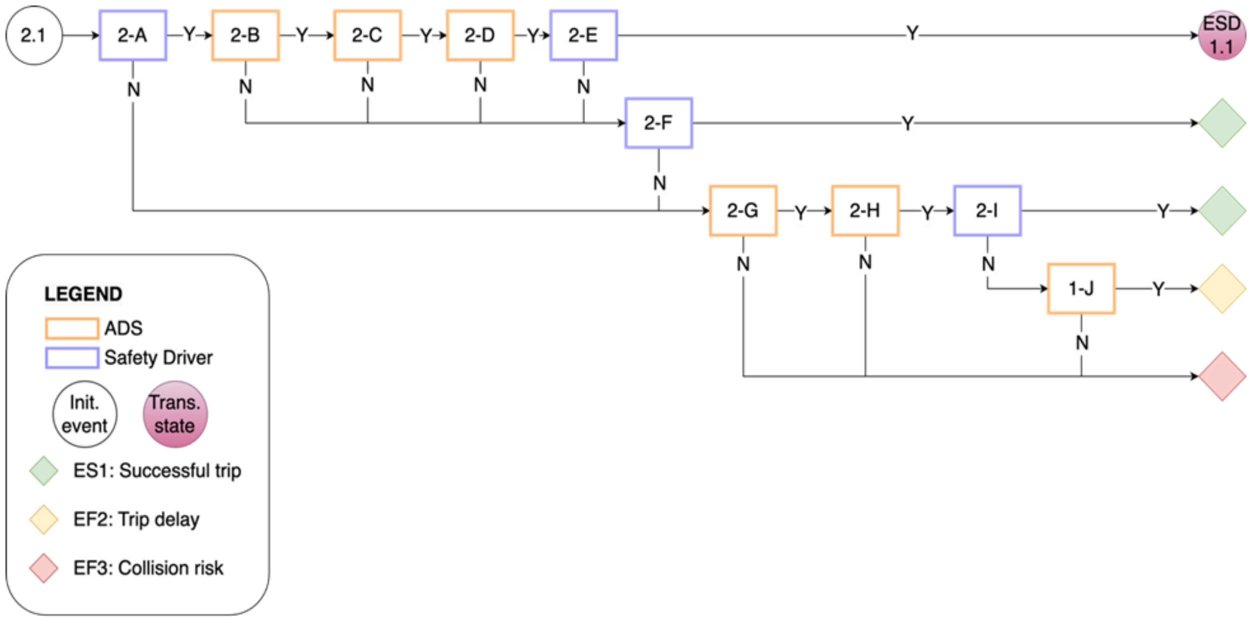


Figure 8. ESD 2.1 – Manual Driving

If it is deemed safe (i.e. there is no imminent collision risk and the vehicle remains in the ODD), the ADS notifies the driver of a handover request. The driver can decide whether to approve the handover request, which then leads to ESD 1.1. If the driver decides to not approve the handover request, they are responsible for completing the DDT, even with the vehicle in the ODD. This may also lead to a successful trip although Automated Driving was not engaged.

Table 9. ESD 2.1 – Event Descriptions

Event	Name	Agent
2.1	The vehicle cleared to begin shift.	-
2-A	Driver correctly performs Manual Driving tasks outside of ODD.	DRI
2-B	ADS correctly detects ODD entrance.	ADS
2-C	ADS determines that it is safe to engage automation.	ADS
2-D	ADS notifies the driver of handover request.	ADS
2-E	Driver requests handover	DRI
2-F	Driver correctly performs Manual Driving task inside ODD	DRI

Event	Name	Agent
2-G	Driver monitoring system detects driver inattention	ADS
2-H	Driver monitoring system alerts the driver of inattention.	ADS
2-I	Driver responds to alert and is able to complete DDT.	DRI
1-J	The vehicle can implement a failure mitigation strategy.	ADS
ESD 1.1	Driver requests handover.	DRI

If the safety driver does not correctly perform the dynamic driving task, whether this occurs inside or outside of the operational design domain, the driver monitoring system, a component of the ADS, nominally detects and alerts the driver of inattention. If the driver responds to the driver monitoring system alert and can complete the driving task, then this is denoted as a successful trip. Alternately, the vehicle can implement a failure mitigation strategy in the case of driver inattention, leading to a trip delay. If the vehicle is unable to successfully implement a failure mitigation strategy, this can lead to a vehicle collision risk.

The ESD 2.2, pictured in **Figure 9** and described in **Table 10** starts with Automated Driving engaged, and the ADS vehicle approaching the limits of the ODD. The ADS is expected to detect the ODD limit approach and notify the safety driver. The driver may also preventatively respond to DDT fallback triggers, independent of whether the ADS alerts the driver. The safety driver could choose to take over, leading to ESD 1.2. If the driver does not detect the limit approach, the ADS is expected to implement a failure mitigation strategy, which involves allowing the vehicle to come to a safe stop, leading to a trip delay. Hence, the safety barriers are constructed hierarchically, with the first being ADS detection, then driver detection and intervention, and finally implementing a failure mitigation strategy.

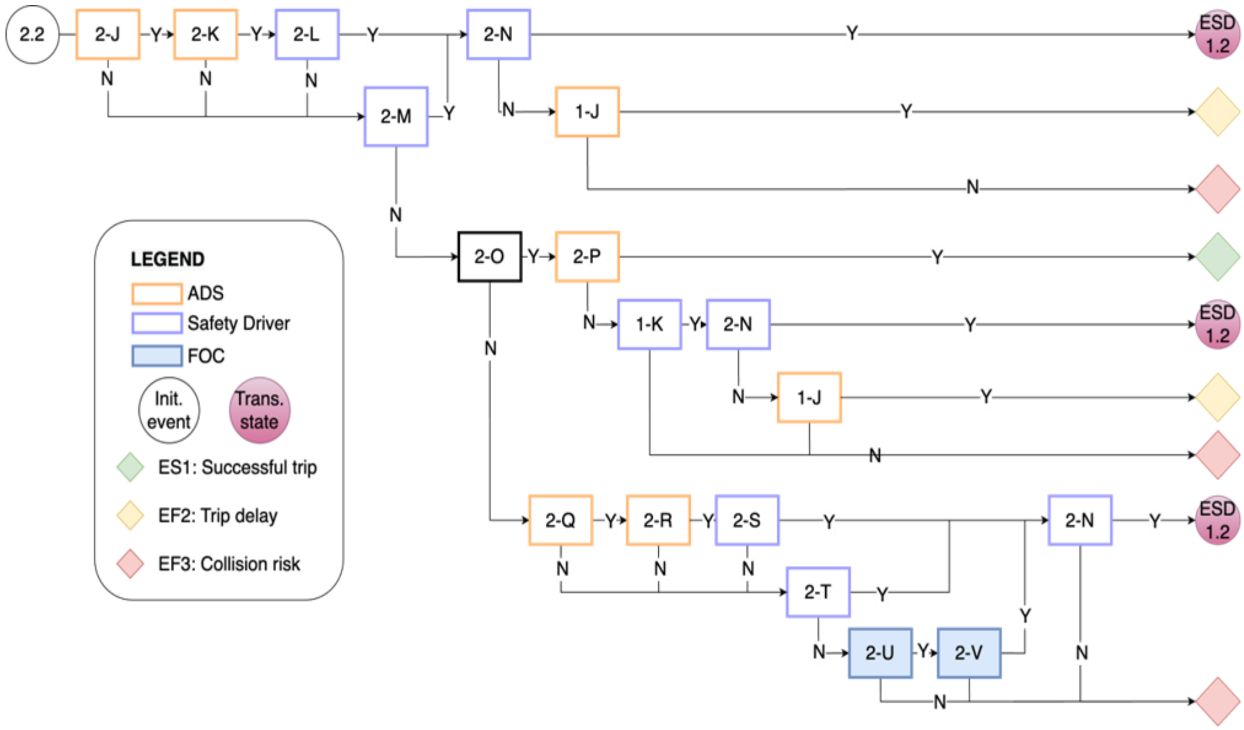


Figure 9. ESD 2.2 – Vehicle Approaches Limits of ODD

If the vehicle exits the ODD, the ADS is expected to alert the safety driver. Even if the ADS does not detect or alert the driver of the ODD breach, the safety driver may independently detect and respond to the ODD breach. If neither the safety driver nor the ADS detect the ODD breach, and the FOC receives a breach notification, the FOC can notify the driver of the need to perform a takeover. If the FOC does not detect the ODD breach, there is an issue in performing the takeover, or a failure mitigation strategy fails, this may lead to an unmitigated risk of collision.

Table 10. ESD 2.2 – Event Descriptions

Event	Name	Agent
2.2	Vehicle approaches limits of ODD.	-
2-J	ADS detects ODD limit approach.	ADS
2-K	ADS notifies driver that vehicle is approaching ODD limits and requests takeover.	ADS
2-L	Driver detects ODD limit approach and ADS takeover request.	DRI

Event	Name	Agent
2-M	Driver detects ODD limit approach.	DRI
2-N	Driver performs takeover of vehicle.	DRI
1-J	Vehicle can implement a failure mitigation strategy.	ADS
2-O	Vehicle remains in ODD.	-
2-P	ADS is able to perform the entire DDT within ODD.	ADS
1-K	Driver detects that DDT-fallback is needed.	DRI
2-Q	ADS detects ODD limit exit.	ADS
2-R	ADS notifies driver and FOC that the vehicle has breached ODD and requests takeover.	ADS
2-S	Driver detects notification of ODD exit and ADS takeover request.	DRI
2-T	Driver detects ODD limit exit.	DRI
2-U	FOC detects ODD limit exit.	FOC
2-V	FOC notifies the driver that the vehicle has exited ODD.	FOC
ESD 1.2	Driver performs take-over.	DRI

Step 4: Concurrent Task Models

The fourth step involves modelling agents' tasks and interactions through CoTA. CoTA is a technique which decomposes high-level goals of each agent into a series of tasks and subtasks to analyze interactions between agents and how they relate to overall goal success. It decomposes tasks based on hierarchical task analysis (HTA), where tasks are re-described until fundamental tasks relating to the interactions between agents appear (Ramos et al. 2020b). These tasks are categorized based on the IDA (information, decision, action) cognitive model (Chang and Mosleh 2007).

IDA was initially developed for human agents, where tasks are decomposed into receiving information from a system (I), planning a course of action (D), and performing the action (A). It has been expanded to provide further details regarding team dynamics (Azarkhil, Mosleh, and Ramos 2025), to study autonomous ships in the maritime industry (Ramos et al. 2020a), and identify factors influencing human-system interactions in automated driving contexts (Correa-Jullian et al. 2024a; Correa-Jullian et al. 2024b). Each task and subtask in CoTA is categorized into a series of plans that denote the order in which the tasks are carried out. The task categories that make up these plans are summarized in **Table 11**.

Table 11. CoTA Task Plan Categories

Task Type	Symbol	Meaning
Parallel	1 // 2	Perform Tasks 1 and 2 at all times
Sequential	1 -> 2	Perform Task 1 then Task 2
Triggered	1 -- 2	Perform Task 2 only if triggered by Task 1
Exclusive	1 or 2	Perform Task 1 or Task 2

CoTA models were developed for the safety driver, ADS, and FOC agents in this system. For illustration purposes, simplified versions of the safety driver (DRI) and ADS CoTAs are presented in **Figure 10** and **Figure 11**, but as needed, connections to other agent CoTAs are shown. The complete FOC CoTA is shown in **Figure 12**. Relevant CoTA subtasks used for metric creation are also shown in the Appendix.

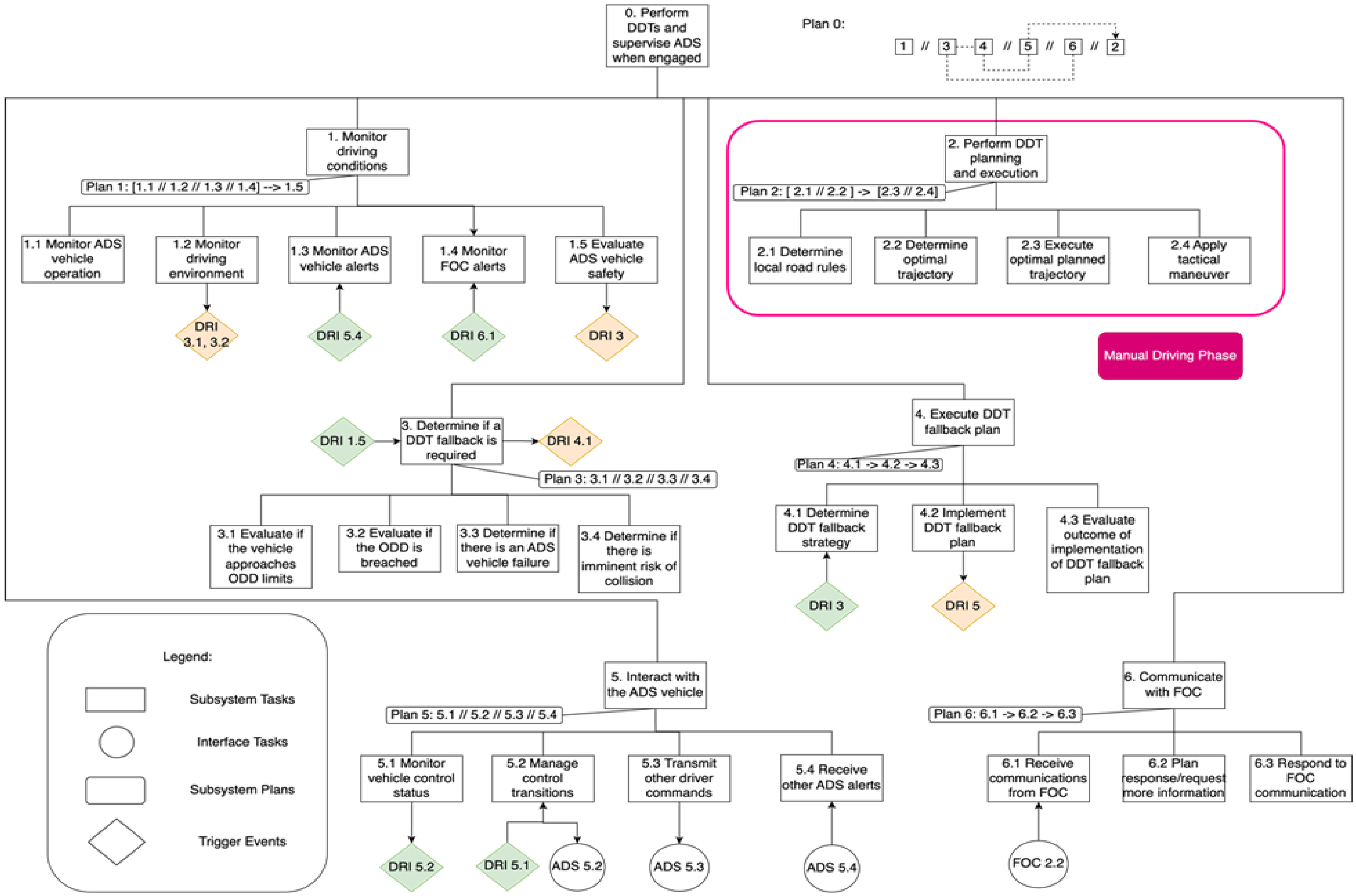


Figure 10. Simplified Safety Driver CoTA Model

As depicted in **Figure 10**, the role of the safety driver in the HD-AV was decomposed into six high-level tasks described in **Table 12**. These tasks involve continuous monitoring of driving conditions (Task 1) and communicating with the FOC (Task 6) when required. During the Manual Driving operational phase, the safety driver is responsible for performing DDT planning and execution (Task 2). Even while the ADS is engaged, the safety driver is expected to use the information from Task 1 to determine whether a DDT fallback is needed (Task 3). If a DDT fallback is needed, Task 3 triggers Task 4, executing the DDT fallback plan. Throughout all operational phases, the safety driver also interacts with the ADS vehicle (Task 5).

The driver monitors the control status, can request transitions and receive transition requests from the ADS, and can transmit and receive ADS alerts. The safety driver's Task 5 interfaces with ADS's Task 5, where the ADS also monitors control status, can request control transitions and receive handover requests, and transmits and receives driver alerts.

Table 12. Safety Driver CoTA Tasks

Num.	Subtask	Type	Description
1	Monitor Driving Conditions	Parallel	The safety driver performs monitoring tasks during all phases of operation. This involves monitoring the ADS vehicle operation, driving environment, alerts from the ADS vehicle, and communications from the FOC. Information gathered from this task supports the other tasks.
2	Perform DDT Planning and Execution	Triggered	This occurs during the Manual Driving phase or is triggered by Task 5, when a control transition occurs to transfer DDT control to the safety driver. When triggered, the driver uses the information from task 1 to fully plan and execute the DDT. The DDT involves employing object and event detection and reaction (OEDR) functions, following local traffic rules, and, if necessary, implementing tactical maneuvers.
3	Determine if a DDT Fallback is Required	Parallel/ Trigger	At all operational phases, the safety driver determines if the situation requires a DDT fallback plan. A DDT fallback plan can be triggered by an ODD breach or limit approach, a vehicle or sensor failure, or by a perceived risk of collision.

Num.	Subtask	Type	Description
4	Execute DDT Fallback Plan	Sequential/ Triggered	This task is triggered by Task 3. The driver determines the DDT fallback strategy and implements a fallback plan. The strategy requires the driver to assess the vehicle condition and determine what the end state should be, either allowing the ADS to continue performing the DDT, requesting a control transition, or requesting an emergency stop. Once planned, the driver implements the fallback plan and evaluates the outcome.
5	Interact with the ADS Vehicle	Parallel	The driver continuously receives and transmits commands to the ADS regarding vehicle control transitions, emergency stop requests, and navigational inputs. Additionally, the driver manages the vehicle control transitions. The driver may request control transitions, i.e., driver-initiated handovers or takeovers and respond to system-initiated requests.
6	Communicate with FOC	Parallel	At all operational phases, the safety driver communicates with the fleet operations center. For this task, the safety driver receives communications from the FOC, plans a response, and then responds to the FOC.

As depicted in **Figure 11**, the role of the ADS in HD-AV operations was decomposed into seven high-level tasks described in **Table 13**. These tasks involve continuous analyzing sensor data to observe the driving environment (Task 1) and performing self-diagnostic tasks to observe ADS operation (Task 6). During the Automated Driving operational phase, the ADS is responsible for performing DDT planning and execution (Task 2), determining whether a DDT fallback is needed (Task 3), and if needed, executing the DDT fallback plan (Task 4). Throughout all operational phases, the ADS also interacts with the safety driver (Task 5). The ADS transmits and receives alerts to the driver regarding control transition events, emergency stops, and navigational inputs, directly interfacing with Task 5 of the safety driver. Additionally, the ADS communicates with the safety driver and FOC about ADS and physical vehicle status, also initiating driver monitoring system alerts (Task 7).

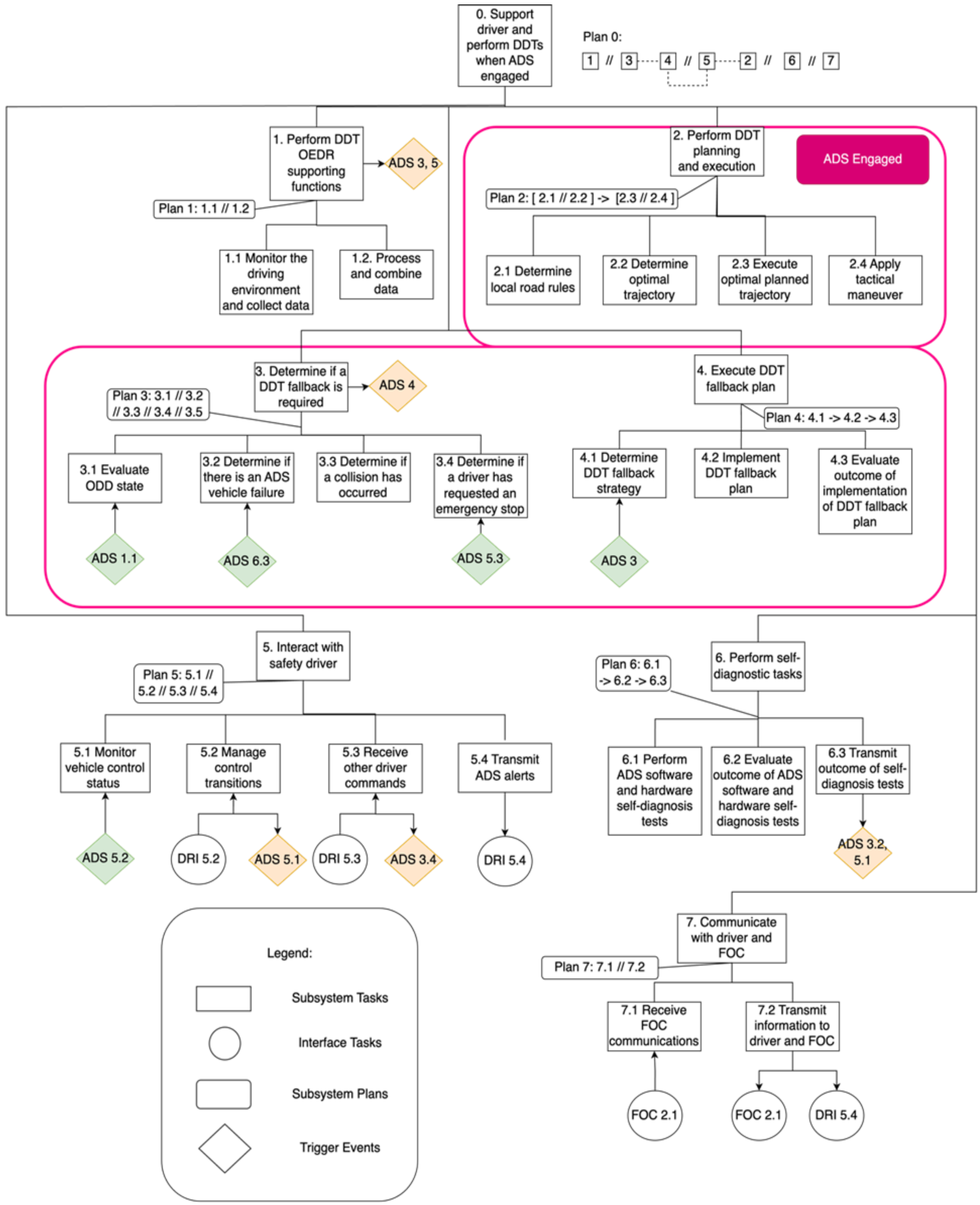


Figure 11. Simplified ADS CoTA Model

Table 13. ADS CoTA Tasks

Num.	Subtask	Type	Description
1	Perform DDT OEDR Supporting Functions	Parallel	The ADS gathers and processes sensor data to gain information about the vehicle, environment, and ODD to support the other parallel tasks.
2	Perform DDT Planning and Execution	Triggered	This task is triggered by Task 5, when a control transition occurs to transfer DDT control to the ADS. The ADS continuously uses information from Task 1 to fully plan and execute the DDT.
3	Determine if a DDT Fallback is Required	Parallel/ Trigger	At all operational phases, the ADS continuously determines whether the situation requires a DDT fallback plan, which can be triggered by an ODD breach or limit approach, a vehicle or sensor failure, a perceived risk of collision, or an emergency stop request initiated by the driver.
4	Execute DDT Fallback Plan	Sequential/ Triggered	This task is triggered by Task 3. The ADS determines the DDT fallback strategy and implements a fallback plan. The strategy can involve continuing the DDT, implementing a minimal risk condition or stable stopped condition. Once planned, the ADS implements the fallback plan and evaluates the outcome.
5	Interact with Safety Driver	Parallel	The ADS continuously receives and transmits commands to the driver regarding vehicle control transitions, emergency stop requests, and navigational inputs.
6	Perform self-Diagnostic Tasks	Parallel	The ADS monitors its subsystems and sensor data to determine if there are any malfunctions in the hardware or software. To do this, it performs self-diagnostic tests that notify the driver.
7	Communicate with Safety Driver and FOC	Parallel	At all operational phases, the ADS alerts the safety driver and the FOC about ADS status, vehicle status, and driver monitoring system alerts.

As depicted in **Figure 12**, the role of the FOC was decomposed into three high-level tasks described in **Table 14**. These tasks involve continuous monitoring of operations (Task 1), including ADS, physical vehicle, and driver monitoring alerts.

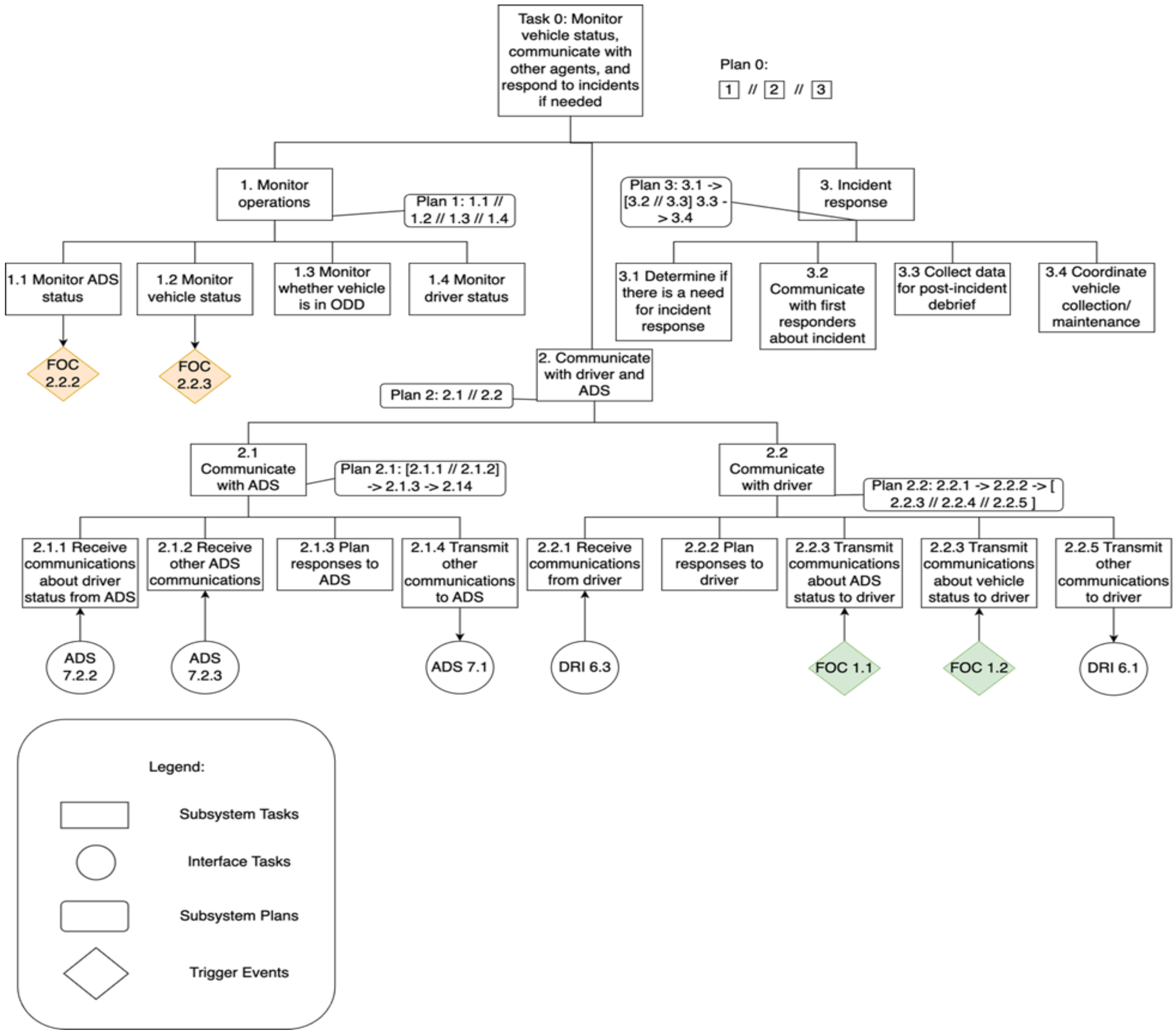


Figure 12. Complete Fleet Operation Center CoTA Model

Additionally, the FOC has access to information about ODD exits so that operators are aware of the physical location and status of the HD-AVs. The next parallel task is communicating with the driver and ADS (Task 2) about status and vehicle alerts. If there are any issues outside of nominal operation, the FOC provides a backup support role and can communicate with the driver about fallback procedures, however the FOC does not provide direct vehicle input or control. Finally, at all stages the FOC assesses the need for and initiates incident response support tasks (Task 3). Again, the FOC does not intervene with the vehicle but coordinates with third party individuals that can provide incident support and coordinate information for a post-incident debrief.

Table 14. FOC CoTA Tasks

Num.	Subtask	Type	Description
1	Monitor Operations	Parallel	The FOC views information about ADS, physical vehicle status, and driver status on a control room dashboard. While this information is limited due to the necessity to monitor multiple vehicles, the FOC receives the information that is relevant.
2	Communicate with Driver and ADS	Parallel	At all operational phases, the FOC communicates with the ADS and the safety driver. The FOC receives driver monitoring system alerts and other communications from the ADS through the dashboard and verbal communications from the safety driver.
3	Incident Response	Parallel/ Triggered	At all operational phases, the FOC continuously determines whether there is a need for incident response, which can be triggered by a vehicle or sensor failure, a perceived risk of collision, an actual collision, or an emergency stop request initiated by the driver. The FOC communicates with first responders and coordinates data collection for the debrief. Additionally, if there is a need to collect the vehicle and transport it back to the maintenance operations center, the FOC coordinates this aspect of the incident response.

Safety-Critical Task Identification

As shown in **Table 12** and **Table 13**, the safety driver and ADS are responsible for similar high-level tasks, such as performing aspects of the DDT planning, execution, and fallback, but they perform these at different operational phases, dictated by control transitions. The safety driver performs DRI Task 2 only during the Manual Driving phase, and the ADS performs ADS Task 2 only during the Automated Driving Engaged phase. Both the safety driver and ADS can assess the need for and implement a DDT fallback at any stage of operation. Task 5 for both DRI and ADS include the control transitions that lead to changes in the operational state of the subsystem. The distinction in driver and ADS DDT fallback is that driver DDT fallback can lead to a takeover request or emergency stop request, but the ADS fallback leads to a minimal risk condition or stable stopped condition, which are initiated by the vehicle and can lead to an operational delay.

To demonstrate the interfacing between separate agent CoTAs for the case of takeover and handover events, the full decomposition of subtask 5.2.1 for the safety driver and ADS is shown in **Figure 13** and **Figure 14** (interface and trigger tasks summarized in **Table 15**). Task 5.2.1 is a subtask of Task 5.2, which denotes the "Manage control transitions" task. This task is divided into managing driver-initiated and system-initiated control transition events. For the case of driver-initiated takeovers, the driver first determines whether a takeover is needed or desired using information from DRI Task 5.1.2, which takes information from DRI Task 1 (driving conditions) and DRI Task 3 (assessing the need for a DDT fallback). If a takeover is desired, the driver performs the takeover (DRI 5.2.1.1.2), and the ADS is required to detect the takeover input (ADS 5.2.1.1.1).

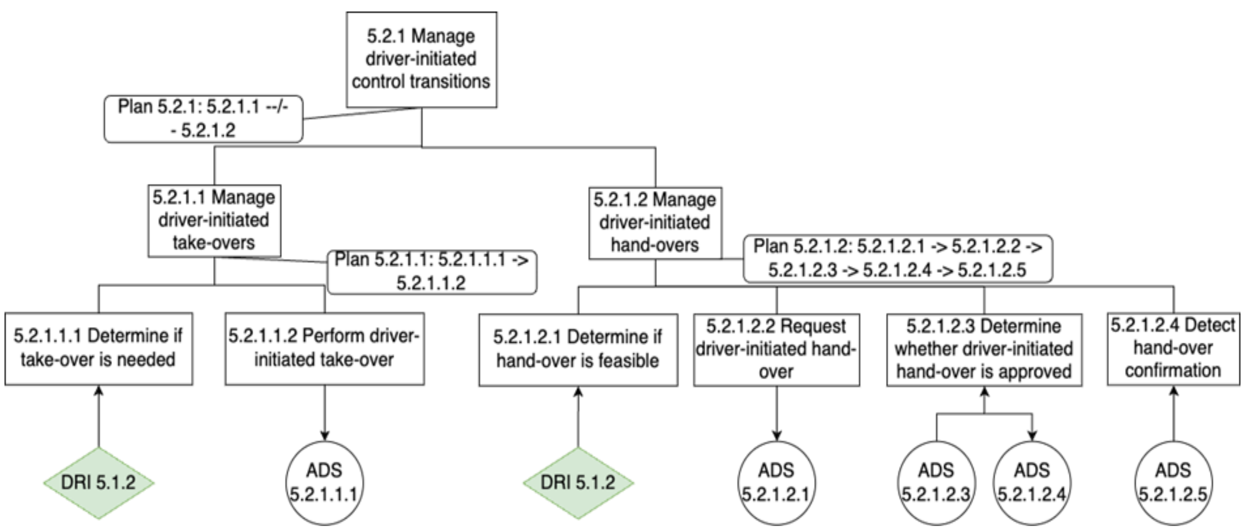


Figure 13. DRI Task 5.2.1 CoTA Decomposition

For the driver-initiated handover, first the driver determines whether an ADS handover is feasible (DRI 5.2.1.2.1) and initiates a request (DRI 5.2.1.2.2). This interfaces with ADS 5.2.1.2.1, which is receiving the handover request, and ADS 5.2.1.2.2, which is the ADS determining whether the handover is feasible. The ADS then communicates the status of the request, which can be an approval or a denial of the request (ADS 5.2.1.2.3). The driver is responsible for determining whether the handover is approved (DRI 5.2.1.2.3), and if it is approved the ADS performs the desired handover (ADS 5.2.1.2.4).

Once the handover occurs, the ADS transmits a confirmation (ADS 5.2.1.2.5), which the driver is responsible for detecting (DRI 5.2.1.2.4). It can be noted that a takeover or handover can be unsuccessful due to a lack of detection, an ADS operation failure, or other issues, and unsuccessful rates of takeovers and handovers are quantified.

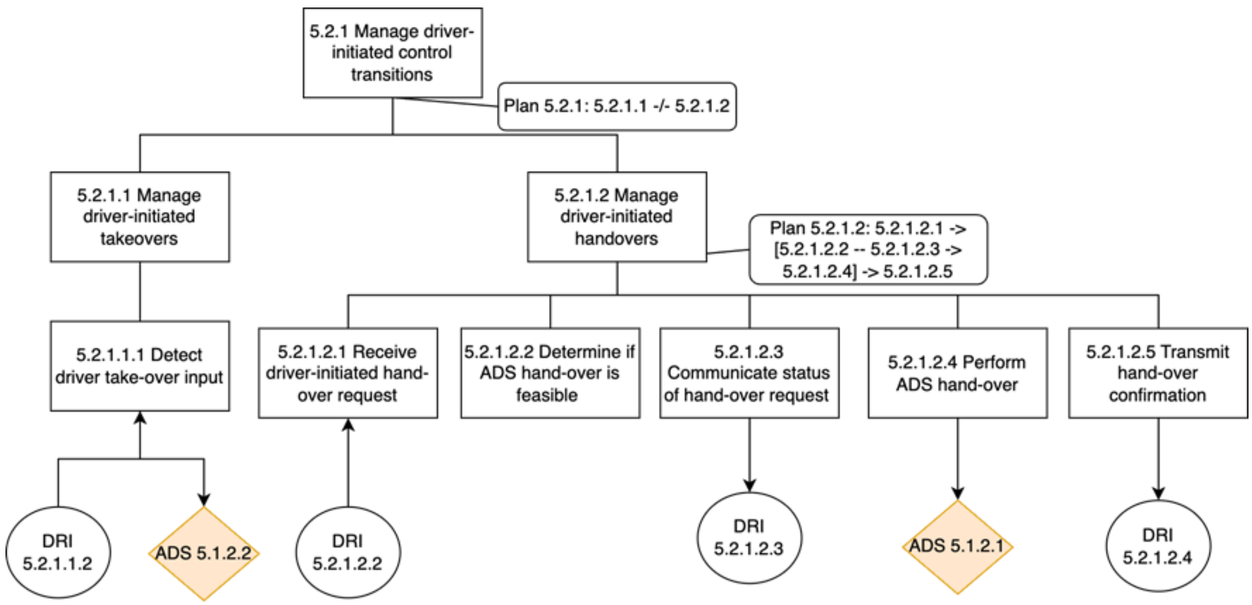


Figure 14. ADS Task 5.2.1 CoTA Decomposition

The control transition can be attempted again after the fact, or if there is a critical issue, the Fallback/minimal risk condition stage or an emergency stop can be triggered. The CoTA provides a structured way to observe agent interactions in the same hierarchy and assess which of these interactions can point to operational risks.

Table 15. Interface Tasks and Triggering Events for DRI/ADS CoTA 5.2.1

Agent	Task Num.	Task
DRI	5.1.2	Determine if vehicle control change is desired
ADS	5.2.1.2.2	Determine if driver is in control of the vehicle
ADS	5.2.1.2.1	Determine if ADS is in control of the vehicle

Derived Safety Metrics

In observing failure paths from the ESD and the tasks that lead to failure modes in the CoTA, a list of human-system interaction metrics was developed and grouped by categories. The metrics were created by determining which factors of the CoTA tasks could be measured to point out potential operational safety weaknesses. The proposed metrics for the modelled HD-AV system are summarized in **Table 16** to **Table 20**. More details are available in the Appendix.

Ideally, these metrics can be tracked in HD-AV simulation and testing to inform design of components like human-system interface and operational tasks in initial stages and validate the system in later testing stages. Additionally, many metrics can be used during nominal operations and are not incident-based, so data trends can be assessed. Most of these metrics would be collected from the ADS data log, which keeps track of sensor and alarm data; operational phase changes, including takeover and handover events; and quality of takeovers based on vehicle dynamics. For qualitative metrics, questionnaires, surveys, and interviews can be employed to determine possible root causes leading to decisions made by the safety drivers and FOC operators assessing the quality of the ADS post-incident response.

As shown in **Table 16**, within the control transitions category, the metric "Rate of Successful Driver-Initiated Takeovers" was determined by observing the interaction between DRI and ADS Task 5.2.1.1, shown in **Figure 13** and **Figure 14**. A failure in detection of driver takeover input would lead to an unsuccessful takeover, which could highlight a potential software or hardware ADS risk. Hence, recording the possible root causes of the driver-initiated takeover can then support system design improvement decisions (e.g., control transition mechanisms) or temporary restrictions in the ODD during operation. Further control transition metrics relate to success rates for other control transition events, and the reasons these control transitions occur. Although the reason for a control transition can be obtained qualitatively through surveys and interviews, by gaining information from initial studies, the potential reasons can be categorized and then tracked in a more systemic manner. In terms of quality of takeover measurements, there has been research into the creation of standardized metrics, three of which are chosen here (Cao et al. 2021).

Table 16. Control Transition Metrics

Num.	Name	Definition	Unit	Data Source	Use Case	Origin
1	Rate of Successful Driver-Initiated Handovers	Ratio of successful driver handovers to total number of handover requests.	%	Data Log	Testing/ Operation	DRI CoTA 5.2.1.2.4, ADS CoTA 5.2.1.2.5, ESD 1.1

Num.	Name	Definition	Unit	Data Source	Use Case	Origin
2	Rate of Successful Driver-Initiated Takeovers	Ratio of successful driver takeovers to total number of takeover attempts.	%	Data Log	Testing/ Operation	DRI CoTA 5.2.1.1.2, ADS CoTA 5.2.1.1.1, ESD 1.2
3	Rate of Successful System-Initiated Handovers	Ratio of successful driver handovers to total number of system-initiated handover requests.	%	Data Log	Testing/ Operation	DRI CoTA 5.2.2.2.3, ADS CoTA 5.2.2.2.5
4	Rate of Successful System-Initiated Takeovers	Ratio of successful driver takeovers to total number of system-initiated takeover requests.	%	Data Log	Testing/ Operation	DRI CoTA 5.2.2.1.2, ADS CoTA 5.2.2.1.3
5	Rate of ADS Handover Approval	Ratio of driver approval of system-initiated handovers to system-initiated handover requests.	%	Data Log	Testing/ Operation	DRI CoTA 5.2.2.2.2
6	Reason for Driver-Initiated Takeover	Category for reason safety driver initiated a takeover (e.g. lack of trust, unnoticed ODD breach).	n/a	Survey or Interview	Testing/ Operation	ESD 1.2, DRI CoTA 5.2.1.1
7	Reason for System-Initiated Takeover	Category for reason ADS initiated a takeover request (e.g. ODD breach, collision risk).	n/a	Data log	Testing/ Operation	ADS CoTA 5.2.2.1.1

Num.	Name	Definition	Unit	Data Source	Use Case	Origin
8	Quality of Takeover (TTC-Based)	Minimum time to collision and maximum resulting lateral and longitudinal acceleration after the initiated takeover request.	sec	Data Log	Testing/ Operation	Literature / System Model
9	Quality of Takeover (Dynamics-Based)	Maximum resulting weighted sum of lateral and longitudinal acceleration after the initiated takeover request.	m/s^ 2	Data Log	Testing/ Operation	Literature / System Model
10	Quality of Takeover (TOT-Based)	Takeover time (TOT) interval between takeover request (TOR) and the driver's first maneuver.	sec	Data Log	Testing/ Operation	Literature / System Model

The alert-related metrics (listed in **Table 17**) were used to measure the ratio of alerts arising from diverse sources (e.g., vehicle-related malfunctions, driver monitoring system, ODD breaches, etc.) as well as the ratio of alerts not responded to by the safety driver. For instance, while it is important to assess whether alerts are detected by the safety driver, this presents significant difficulties. Hence, the metric "Alerts Not Acted On" can serve as a partial indicator that indicate a need for developing improvements being at human-system interface display, alert design, or driver training level.

The incident metrics (listed in **Table 18**) track incident rate per vehicle miles-traveled (VMT), and these rates are classified into incidents with the driver in control and ADS in control. Additionally, they are disaggregated into incident severity levels, with the levels being "Traffic Disruption Only", "Property Damage Only", and "Collision", which includes incidents with damage to other vehicles, fatalities, and injuries. These are all derived from FOC CoTA Task 3, which denotes incident response (See **Figure 12**, **Table 14**).

Table 17. Alert-Related Metrics

Num.	Name	Definition	Unit	Data Source	Use Case	Origin
1	Alerts Resulting from Vehicle-Related Malfunctions	Ratio of alerts coming from vehicle sensor, ADS or vehicle malfunction to total number of alerts.	%	Data Log	Testing/ Operation	DRI CoTA 5.4.2
2	Alerts Resulting from Onboard Safety Driver	Ratio of alerts generated by the driver monitoring system to total number of alerts.	%	Data Log	Testing/ Operation	DRI CoTA 5.4.3
3	Alerts Resulting from Environment	Ratio of alerts coming from road conditions, ODD breach to total number of alerts.	%	Data Log	Testing/ Operation	DRI CoTA 5.4.1
4	Alerts Not Acted On	Ratio of alerts not responded to by safety driver to total number of alerts.	%	Data Log	Testing/ Operation	ESD 2.2, Event 2-I

Table 18. Incident-Related Metrics

Num.	Name	Definition	Unit	Data Source	Use Case	Origin
1	Incident Rate - Driver in Control	Rate of incidents (collision, property damage) leading to an initiation of post-incident procedures while safety driver is in control.	#/VMT	Data Log	Operation	FOC CoTA 3

Num.	Name	Definition	Unit	Data Source	Use Case	Origin
2	Incident Rate - ADS in Control	Rate of incidents (collision, property damage) leading to an initiation of post-incident procedures while ADS is in control.	#/VMT	Data Log	Operation	FOC CoTA 3
3	Incident Rate – Driver in Control – Disruption Only	Rate of incidents while safety driver is in control that result in only a traffic disruption.	#/VMT	Data Log	Operation	FOC CoTA 3
4	Incident Rate – Driver in Control – Property Damage Only	Rate of incidents while safety driver is in control that result in only property damage.	#/VMT	Data Log	Operation	FOC CoTA 3
5	Incident Rate – Driver in Control – Collision	Rate of incidents while safety driver is in control that result in a collision with a pedestrian or vehicle.	#/VMT	Data Log	Operation	FOC CoTA 3
6	Incident Rate – ADS in Control – Disruption Only	Rate of incidents while ADS is in control that result in only a traffic disruption.	#/VMT	Data Log	Operation	FOC CoTA 3
7	Incident Rate – ADS in Control – Property Damage Only	Rate of incidents while ADS is in control that result in only property damage.	#/VMT	Data Log	Operation	FOC CoTA 3

Num.	Name	Definition	Unit	Data Source	Use Case	Origin
8	Incident Rate – ADS in Control – Collision	Rate of incidents while ADS is in control that result in a collision with a pedestrian or vehicle.	#/VMT	Data Log	Operation	FOC CoTA 3

The fallback metrics (listed in **Table 19**) refer to rates of a failure mitigation strategy, emergency stops, and cases of ADS fallback resulting from driver inaction. Additionally, the "Delay Resulting from Fallback" aims to quantify the physical time delay resulting from fallback procedures. The fallback metrics can be used to assess fallback scenarios, even though an incident may not occur and determine the rates at which the ADS and the safety driver initiate fallback procedures.

As listed in **Table 20**, the Human-ADS Trust metrics consist of measurements of disagreement with ADS maneuvers and a measure of the time spent in the Ready to Engage phase without engaging the ADS. These trust metrics can also be supplemented with existing human factors studies on ADS trust (Yang et al. 2018; Payre, Perelló-March, and Birrell 2023; Manchon, Bueno, and Navarro 2022).

While the derived human-system interaction metrics serve as a starting point for observing trends in HD-AV system behavior, they may not be relevant for all HD-AV operational scenarios. Thus, the way they are derived from the ESD scenario modelling and CoTA task decomposition can be used to develop metrics for a more specific HD-AV configuration.

Table 19. Fallback-Related Metrics

Num.	Name	Definition	Unit	Data Source	Use Case	Origin
1	Rate of Emergency Stop Requests	Rate of driver requests an emergency stop (per vehicle miles driven).	#/VMT	Data Log	Testing/ Operation	DRI CoTA 5.3.2
2	Rate of ADS-Initiated Failure Mitigation Strategy	Rate of ADS initiated a failure mitigation strategy being implemented per mile.	#/VMT	Data Log	Testing/ Operation	ADS CoTA 4

Num.	Name	Definition	Unit	Data Source	Use Case	Origin
3	Rate of Safety Driver Manual Fallback	Rate of safety driver takeovers when fallback is needed.	#/VMT	Data Log	Operation	DRI CoTA 4
4	ADS Fallback Due to Driver Inaction	Rate of ADS initiated fallback, specifically when driver monitoring system detects inattention and lack of response.	#/VMT	Data Log	Operation	ADS CoTA 4
5	Delay Resulting from Fallback	Per-shift time delay resulting from implementation of fallback procedures.	hours	Data Log	Testing/Operation	FOC CoTA 3

Table 20. Human-ADS Trust Metrics

Num.	Name	Definition	Unit	Data Source	Use Case	Origin
1	Divergence from ADS Decisions	Ratio of maneuvers conducted by safety driver during Automated Driving phase differing from ADS maneuver intention	%	Data Log	Testing	DRI CoTA 5.1.2
2	Time Spent in Ready to Engage Phase	Time spent in the Manual Driving—Ready to Engage phase without engaging ADS	hours/ VMT	Data Log	Testing/ Operation	System model

The aim in creating these metrics is to implement leading measures that can point to potential hazards, as well as increase a focus on operational safety and trends rather than solely focusing on functional safety. While many metrics have been developed and validated specifically for ADS systems as well as for certain control transition events, there is a lack of implementation of human-system interaction metrics in a systemic way. Arriving at these measures of human-system interaction using task and agent-based models allows for a direct observation of the interactions that occur and how they can affect the overall system safety.

Discussion

Although not all events leading to potential risk can be directly measured, safety metrics can point to contributing factors that can be addressed to manage the system's risk. Further, leading metrics can be measured prior to incidents occurring, providing a proactive view of risk assessment. For instance, while it is important to assess whether alerts are detected by the safety driver, this presents significant difficulties. Hence, the metric “Alerts Not Acted On” can serve as a partial indicator that indicate a need for developing improvements in human-system interface display, alert design, or driver training level.

One of the primary topics that has been assessed in ADS research is the degree of trust that human operators inside and outside the vehicle have on the system (Yang et al. 2018). Although somewhat of a qualitative assessment, trust can be estimated in a numerical sense by employing metrics in addition to qualitative information gleaned from surveys and interviews. Another consideration of human system interaction for heavy duty vehicles involves the team of human operators onboard, if there is both a safety driver and a safety operator. The dynamics between these two operators can affect the decision-making and action-taking processes regarding the ADS. For instance, if a safety operator notes a potential collision risk or need for takeover, but the driver ignores them and is reliant on either their own perspective or that of the ADS sensors, these disagreements may develop into more complex hazard scenarios. Although the disagreements between safety driver, safety operator, and ADS can be observed in a numerical manner denoting number of times a disagreement occurs, it is likely that a more complete picture will be formed through incorporating a post-shift debrief, in which questionnaires will assess the causes of such a disagreement.

Model-based approaches to metric creation can allow for analysis of low-level tasks that can point to areas for system safety improvement. A combined approach relying on model-based risk assessment and benchmarked simulators, such as CARLA or OpenCDA, can lead to overall improvements prior to the system development and implementation (Dosovitskiy et al. 2017; Xu et al. 2021; 2023). The advantage of combining simulation- and model-based approaches lies in the systematic methods available to model the system's hardware, software, and human elements. Model-based approaches provide traceability as opposed to purely data-driven metrics derived directly from simulations or testing, with the added benefit of being able to be integrated into early design stages, and evolving during system development, certification, and operational phases. Indeed, developing quality safety metrics can play a significant role when assessing operational safety at later stages of system deployments, assessing their evolution over extended periods of time or miles driven.

The autonomy level of current commercial HD-AV systems is commonly advertised at an SAE Level 4, which is defined as the ADS performing the entire DDT and all associated fallback procedures. However, in current testing scenarios, there is always a safety driver in the vehicle who can intervene in case of emergencies. The active role of the safety driver is therefore assumed to be a long-term and ongoing part of ADS testing and preparation for deploying and possibly upgrading their ADSs. Thus, for those operators for whom a safety driver will remain a constant presence, the effective level of autonomy may decrease to Level 3 according to current J3016 definitions; however, since the ADS nominally plans and implements the DDT fallbacks, further discussions may be required to assess intermediate levels of automation. There has been criticism of the currently defined SAE Levels, especially that their strictly linear progression overlooks the necessary hybridity of interactions between human and automated technologies (Hopkins and Schwanen 2021). The HD-AV system can be referred to as existing at a nominal Level 4 of automation; namely it has characteristics of a Level 3 system, however at a nominal state it can perform all the tasks expected of a Level 4 system. As HD-AV companies more clearly define roles and tasks, the autonomy state and the presence of safety drivers and operators can be reassessed.

Since the interplay between safety drivers and the ADS plays a large role in the division of tasks and operation of heavy-duty systems, there is a need to include safety metrics about how these humans and systems interact. This can help inform the designs of ADS systems in turn; for instance, making human-system interfaces more effective for safety drivers, developing more robust driver monitoring systems, and designing alert functions. In addition, observing these metrics can help with designing HD-AV operations for both efficiency and safety without promoting one at the expense of the other.

A focus on human-centered metrics has been applied in various industries such as nuclear and maritime, to develop effective human reliability assessments which are conducted through models such as Phoenix, SPAR-H, and IDAC (Chang and Mosleh 2007; Ekanem, Mosleh, and Shen 2016). Existing human reliability assessment methods can be expanded upon to incorporate automated vehicles in their models and potentially be used to expand current probabilistic risk assessments conducted to inform the design and regulation of these automated driving systems. Additionally, many industries such as occupational health and safety and commercial aviation have begun to prioritize leading metrics as a proactive view of operational safety (Bayramova et al. 2023; Sheehan et al. 2016). A combination of leading metrics and human-system interaction metrics represents a necessary push to operational safety for novel automated driving technologies and should be implemented in the upcoming HD-AV industry.

In summary, due to regulatory barriers, legal liability, and the need to reduce risks, safety drivers will likely continue to be involved in some HD-AV operations in various settings beyond testing. Therefore, it is necessary to observe the interactions between the human and machine agents in this system to assess their safety. Using a comprehensive set of human-system interaction metrics will inform operational and system design during preliminary testing phases. These metrics can improve the design of components such as human-system interface in the vehicle and inform the design of operational tasks. Additionally, metrics can assess trends or point to needed changes during road testing stages and eventual public deployment. This report demonstrates the use of a framework leveraging ESD and CoTA models to derive human-safety interaction metrics that can inform design and development of safe HD-AV systems. Future work can be done towards incorporating other methods, such as System-Theoretic Process Analysis (STPA), to provide alternative characterization of the HD-AV systems, leading to a more comprehensive hazard identification analysis and surrogate safety metric construction (Correa-Jullian et al. 2024c; 2024a). The metric derivation framework presented here can be adapted by HD-AV fleet operators to reflect the respective company's specific operational scenarios and agent tasks. Additionally, the evolution of selected metrics can be observed over time or miles driven in operation in order to make decisions and assess changes made.

Concluding Remarks

With the increased focus on HD-AV development, testing, and deployment, ensuring the safety of human-system interaction remains a central challenge. While many long-term operations envision fully driverless applications at Level 4 ADS, in the short-to-medium term, many use cases will continue to require onboard and remote human supervision. Automation introduces several human factors and reliability concerns, including complacency, loss of situational awareness, increased workload during control transitions, distraction, and skill degradation. These challenges are well-documented across high-risk domains and are especially relevant to HD-AV operations, where the division of tasks between automation and human agents varies by context and continues to evolve. Despite these challenges, safety drivers and remote operators serve as critical safeguards during system failures or unexpected conditions. Understanding the design and operational constraints that ensure their presence effectively reduces risk is fundamental. Their roles must be clearly defined, supported by effective interfaces, and grounded in system designs that reflect human capabilities and limitations, enabling timely human intervention when needed.

Developing human-system interaction metrics is a foundational step toward this goal. These metrics allow researchers, designers, and regulators to track indicators of safety performance related to human involvement and understand how humans respond to system behavior, alerts, control transitions, and anomalies. Such metrics are particularly important in the HD-AV context, where operational models remain fluid and vary in their integration of safety drivers, operators, and remote supervision. Without standardized architectures or protocols, it becomes even more important to assess safety trends using structured, context-sensitive measures. This work demonstrates a framework for deriving such metrics using risk-informed modeling approaches, including Event Sequence Diagrams (ESD) and Concurrent Task Analysis (CoTA). These methods enable systematic identification of key human-machine interface points and help isolate tasks or control transitions that are most likely to introduce safety risks. The resulting metrics go beyond system-centric measures and incorporate factors like alert responsiveness, trust dynamics, takeover quality, and the roles of multiple human agents in shared control.

As HD-AV concepts and regulations mature, a forward-looking safety strategy must include leading and interaction-focused metrics—not just lagging, incident-based indicators. These metrics should evolve alongside the system, from early design through simulation, testing, deployment, and post-deployment evaluation. They can inform iterative system and operational design by highlighting failure points or degraded performance before incidents occur. Doing so will help ensure that human oversight remains effective and that automation is deployed in a way that enhances safety. As the industry progresses, these metrics can serve as a practical foundation for improving transparency, supporting operator training, refining human-machine interfaces, and enabling safer deployment of HD-AV technologies at scale.

Appendix

The appendix includes the control transition ESDs representing the takeover and handover events, and CoTA model diagrams relevant for the creation of the derived human-system safety metrics.

Appendix A: Event Sequence Diagrams

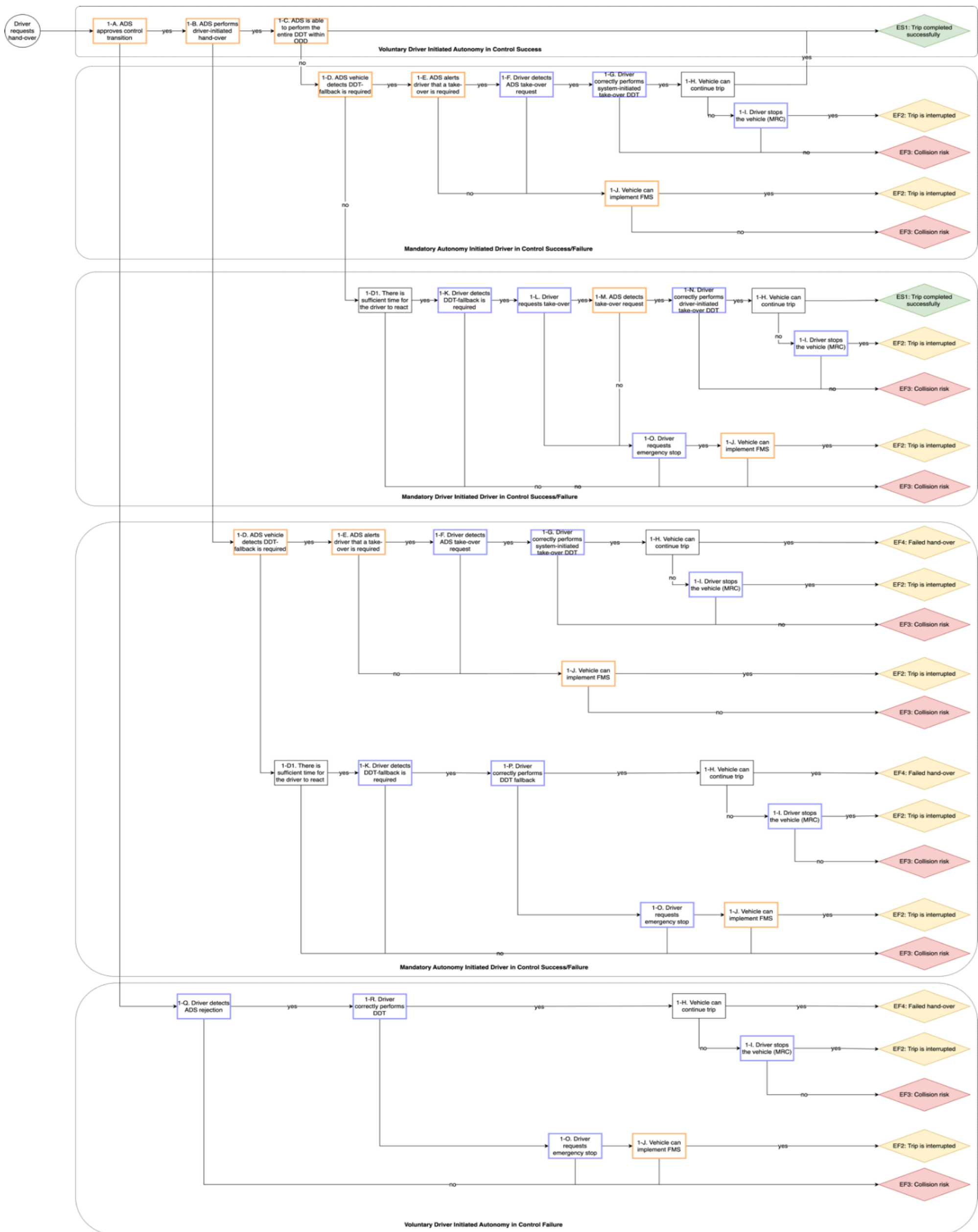


Figure 15. ESD 1.1 – Driver Requests Handover

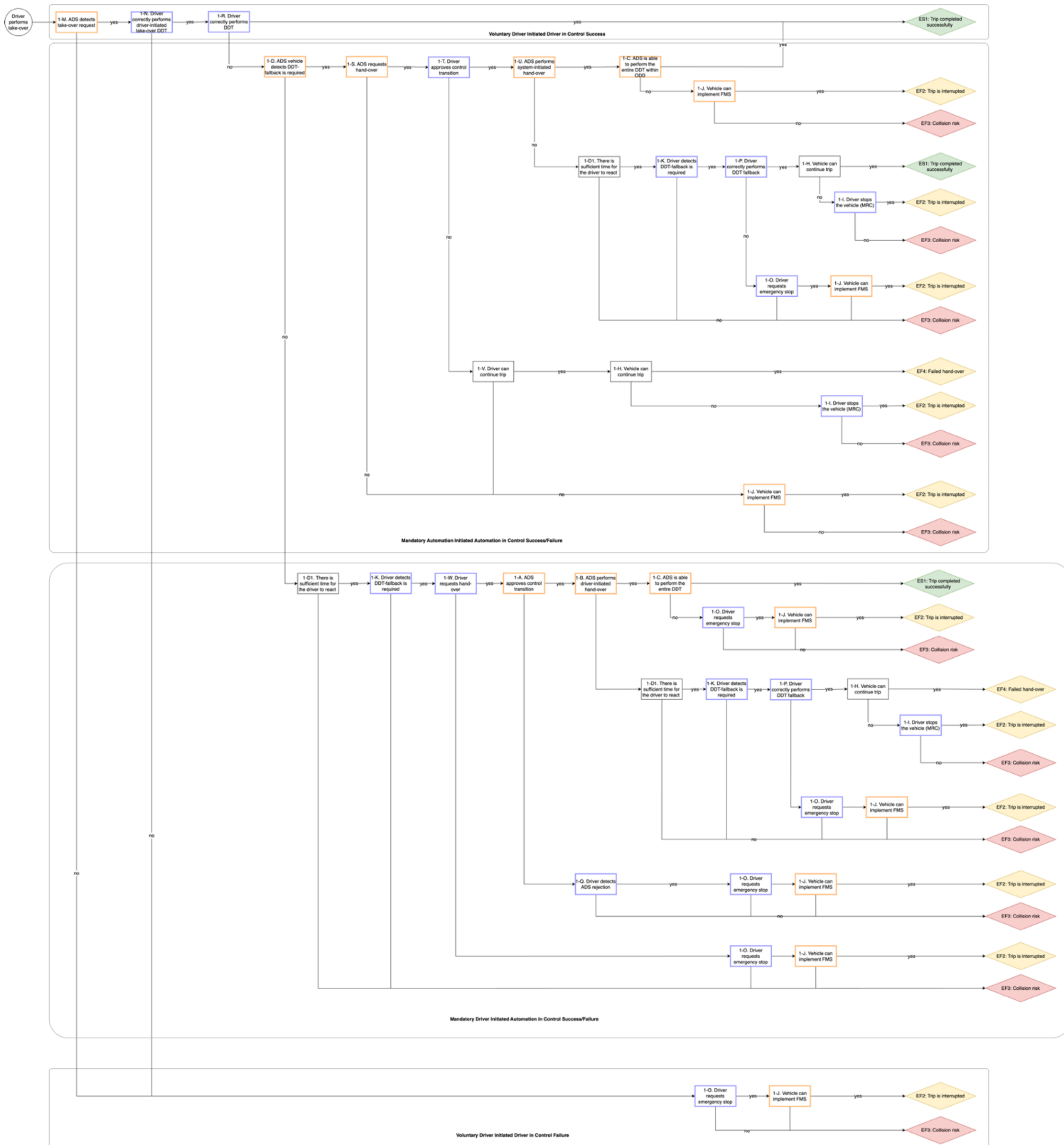


Figure 16. ESD 1.2 – Driver Performs Takeover

Appendix B: Concurrent Task Analysis Diagrams

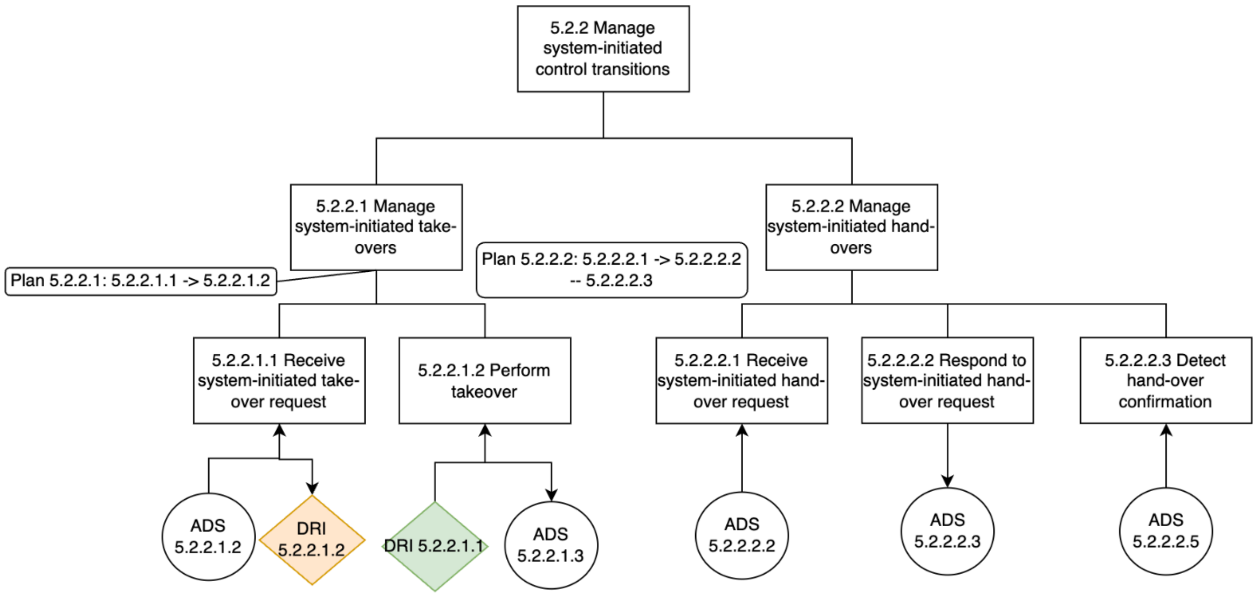


Figure 17. DRI 5.2.2 CoTA

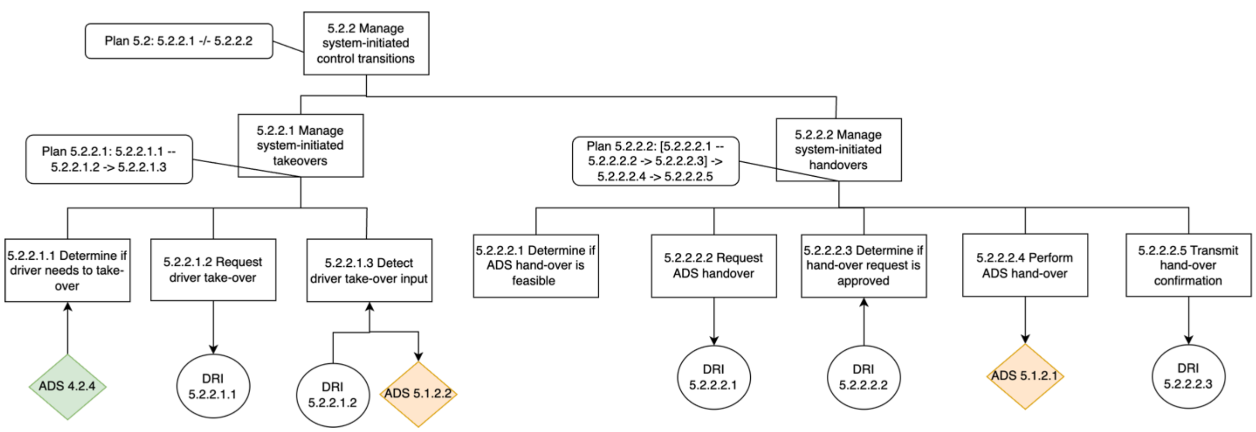
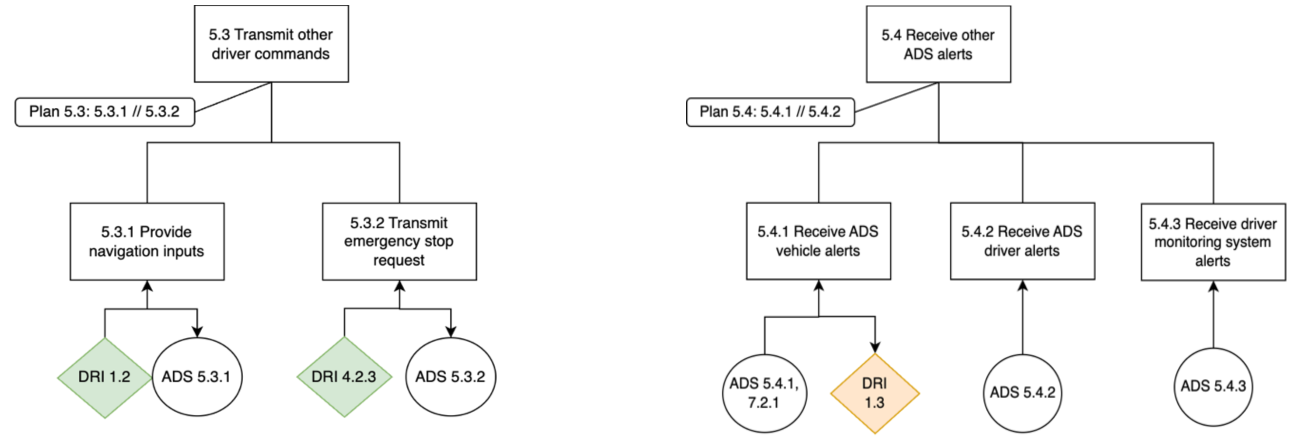


Figure 18. ADS 5.2.2 CoTA



(a) DRI Task 5.3

(b) DRI Task 5.4

Figure 19. DRI Tasks (a) 5.3 and (b) 5.4 CoTA

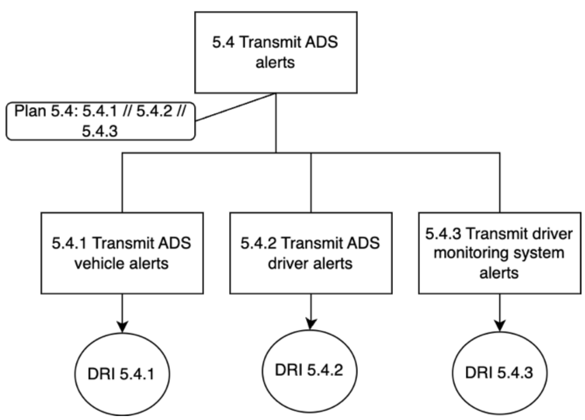


Figure 20. ADS 5.4 CoTA

References

Aurora. 2021. “Voluntary Safety Self-Assessment.” https://aurora.tech/safety.

Automated Vehicle Safety Consortium. 2019. “AVSC Best Practice for In-Vehicle Fallback Test Driver Selection, Training, and Oversight Procedures for Automated Vehicles Under Test.” *SAE Industry Technologies Consortium*. https://www.sae.org/standards/content/avsc00001201911/.

———. 2021. “AVSC Best Practice for Evaluation of Behavioral Competencies for Automated Driving System Dedicated Vehicles (ADS-DVs).” *SAE Industry Technologies Consortium*.

———. 2023. “AVSC Best Practice for ADS Remote Assistance Use Case.” *SAE Industry Technologies Consortia*.

Ayas, Suzan, Birsen Donmez, Xing Tang, and Corresponding Author. 2023. “Drowsiness Mitigation Through Driver State Monitoring Systems: A Scoping Review.” *Human Factors: The Journal of the Human Factors and Ergonomics Society* 0 (0): 1–26. https://doi.org/10.1177/00187208231208523.

Azarkhil, Mandana, Ali Mosleh, and Marilia Ramos. 2025. “Team-Centered IDAC: Modeling and Simulation of Operating Crew in Complex Systems – Part 1: Team Model and Fundamentals.” *Reliability Engineering & System Safety* 254 (February):110541. https://www.sciencedirect.com/science/article/pii/S0951832024006136?via%3Dihub.

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.

Bhoopalam, Anirudh Kishore, Roy van den Berg, Niels Agatz, and Caspar G. Chorus. 2023. “The Long Road to Automated Trucking: Insights from Driver Focus Groups.” *Transportation Research Part C: Emerging Technologies* 156 (November):104351. https://doi.org/10.1016/J.TRC.2023.104351.

Blanco, Myra, Michelle Chaka, L Stowe, H Clay Gabler, Kenneth Weinstein, Ronald B Gibbons, Luke Neurauter, et al. 2020. “FMVSS Considerations for Vehicles With Automated Driving Systems: Volume 1.” Edited by Virginia Tech Transportation Institute. https://rosap.ntl.bts.gov/view/dot/54287.

Bogg, Adam, and Stewart Birrell. 2025. “Overloaded, Underloaded or in Control: How Many Automated Vehicles Can One Person Supervise?” *Computers in Human Behavior* 170 (September):108690. https://doi.org/10.1016/J.CHB.2025.108690.

Bonnefon, Jean-François. 2021. *The Car That Knew Too Much: Can a Machine Be Moral?* MIT Press.

Cao, Yining, Feng Zhou, Elizabeth M. Pulver, Lisa J. Molnar, Lionel P. Robert, Dawn M. Tilbury, and X. Jessie Yang. 2021. “Towards Standardized Metrics for Measuring Takeover Performance in Conditionally Automated Driving: A Systematic Review.” *Proceedings of the Human Factors and*

Ergonomics Society 65 (1): 1065–69. https://doi.org/10.1177/1071181321651213;REQUESTEDJOURNAL:JOURNAL:PROE;WGROUP:STRING:PUBLICATION.

Castro, Candida, P. Pablo Doncel, Rubén D. Ledesma, Silvana A. Montes, D. Daniela Barragan, Oscar Oviedo-Trespalacios, Alessandra Bianchi, Natalia Kauer, Weina Qu, and Jose Luis Padilla. 2024. "Measurement Invariance of the Driving Inattention Scale (ARDES) across 7 Countries." *Accident Analysis and Prevention* 195 (February):107412. https://doi.org/10.1016/j.aap.2023.107412.

Castro, Candida, Jose Luis Padilla, Pablo Doncel, Pedro Garcia-Fernandez, Petya Ventsislavova, Eduardo Eisman, and David Crundall. 2019. "How Are Distractibility and Hazard Prediction in Driving Related? Role of Driving Experience as Moderating Factor." *Applied Ergonomics* 81 (November):102886. https://doi.org/10.1016/J.APERGO.2019.102886.

Chaka, Michelle, Myra Blanco, Loren Stowe, Joshua McNeil, Kevin Kefauver, Vikki L Fitchett, Kaitlyn E Fitzgerald, et al. 2021. "FMVSS Considerations for Vehicles With Automated Driving Systems: Volume 2" 1 (April): 630p. https://rosap.ntl.bts.gov/view/dot/54442%0Ahttps://trid.trb.org/view/1768673.

Chang, Y.H.J., and A. Mosleh. 2007. "Cognitive Modeling and Dynamic Probabilistic Simulation of Operating Crew Response to Complex System Accidents. Part 1: Overview of the IDAC Model." *Reliability Engineering and System Safety* 92 (8): 997–1013. https://doi.org/10.1016/j.ress.2006.05.014.

Chu, Mengdi, Keyu Zong, Xin Shu, Jiangtao Gong, Zicong Lu, Kaimin Guo, Xinyi Dai, and Guyue Zhou. 2023. "Work with AI and Work for AI: Autonomous Vehicle Safety Drivers' Lived Experiences." *ArXiv Preprint ArXiv …* 1 (1). https://doi.org/10.1145/3544548.3581564.

Correa-Jullian, Camila, John McCullough, Marilia Ramos, Ali Mosleh, and Jiaqi Ma. 2023. "Safety Hazard Identification of Inspection and Maintenance Operations for Automated Driving Systems in Mobility as a Service." In *Proceeding of the 33rd European Safety and Reliability Conference*, 281–88. Singapore: Research Publishing Services. https://doi.org/10.3850/978-981-18-8071-1_P326-cd.

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. "Exploring Human-Autonomy Teams in Automated Driving System Operations." In *2024 IEEE 4th International Conference on Human-Machine Systems (ICHMS)*, 1–6. IEEE. https://doi.org/10.1109/ICHMS59971.2024.10555762.

———. 2024c. "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.

Correa-Jullian, Camila, Marilia Ramos, Ali Mosleh, and Jiaqi Ma. 2024a. "Exploring Performance Shaping Factors for Human-Autonomy Teams in Automated Driving System Operations." In *17th International Conference on Probabilistic Safety Assessment and Management & Asian Symposium on Risk Assessment and Management (PSAM17&ASRAM2024).*

———. 2024b. "Task Allocation and Control Transitions in Autonomous Driving System Operations." In *34th European Safety and Reliability Conference (ESREL 2024), Cracow, Poland, 2024*. Vol. 1.

Cosmin-Spanoche, Anna, Camila Correa-Jullian, Xin Xia, Ali Mosleh, and Jiaqi Ma. 2024. "Exploring Safety-Related Metrics to Assess Human-System Interactions in Heavy-Duty Automated Vehicles." In *34th European Safety and Reliability Conference (ESREL 2024), Cracow, Poland, 2024*. Vol. 5.

Cosmin-Spanoche, Anna, Camila Correa-Jullian, Xin Xia, Ali Mosleh, and Jiaqi Ma. 2024. "Deriving Human-System Safety Metrics for HD-AV Applications through Concurrent Task Analysis." In *17th International Conference on Probabilistic Safety Assessment and Management & Asian Symposium on Risk Assessment and Management (PSAM17&ASRAM2024).*

DeGuzman, Chelsea A., Dina Kanaan, Samantha A. Hopkins, and Birsen Donmez. 2021. "Takeover Request (TOR) Effects during Different Automated Vehicle Failures." *Journal of Intelligent Transportation Systems*, March, 1–31. https://doi.org/10.1080/15472450.2021.1891536.

Dong, Yanchao, Zhencheng Hu, Keiichi Uchimura, and Nobuki Murayama. 2011. "Driver Inattention Monitoring System for Intelligent Vehicles: A Review." *IEEE Transactions on Intelligent Transportation Systems* 12 (2): 596–614. https://doi.org/10.1109/TITS.2010.2092770.

Dosovitskiy, Alexey, German Ros, Felipe Codevilla, Antonio Lopez, and Vladlen Koltun. 2017. "CARLA: An Open Urban Driving Simulator," no. CoRL, 1–16. http://arxiv.org/abs/1711.03938.

Ekanem, Nsimah J., Ali Mosleh, and Song Hua Shen. 2016. "Phoenix - A Model-Based Human Reliability Analysis Methodology: Qualitative Analysis Procedure." *Reliability Engineering and System Safety* 145 (January):301–15. https://doi.org/10.1016/j.ress.2015.07.009.

Figalová, Nikol, Hans Joachim Bieg, Michael Schulz, Jürgen Pichen, Martin Baumann, Lewis L. Chuang, and Olga Pollatos. 2023. "Fatigue and Mental Underload Further Pronounced in L3 Conditionally Automated Driving: Results from an EEG Experiment on a Test Track." In *International Conference on Intelligent User Interfaces, Proceedings IUI*, 64–67. Association for Computing Machinery. https://doi.org/10.1145/3581754.3584133.

Hopkins, Debbie, and Tim Schwanen. 2021. "Talking about Automated Vehicles: What Do Levels of Automation Do?" *Technology in Society* 64 (October 2020): 101488. https://doi.org/10.1016/j.techsoc.2020.101488.

Horberry, Tim, Christine Mulvihill, Michael Fitzharris, Brendan Lawrence, Mike Lenne, Jonny Kuo, and Darren Wood. 2022. "Human-Centered Design for an In-Vehicle Truck Driver Fatigue and Distraction Warning System." *IEEE Transactions on Intelligent Transportation Systems* 23 (6): 5350–59. https://doi.org/10.1109/TITS.2021.3053096.

International Organization for Standardization. 2018. “ISO 26262:2018, Road Vehicles — Functional Safety.” 2018. https://www.iso.org/standard/68383.html.

———. 2022. “ISO 21448:2022, Road Vehicles — Safety of the Intended Functionality.” 2022. https://www.iso.org/standard/77490.html.

Khan, Muhammad Qasim, and Sukhan Lee. 2019. “A Comprehensive Survey of Driving Monitoring and Assistance Systems.” *Sensors 2019, Vol. 19, Page 2574* 19 (11): 2574. https://doi.org/10.3390/S19112574.

Kodiak.AI. 2020. “Kodiak Safety Report.” https://kodiak.ai/safety-report.

Koesdwiady, Arief, Ridha Soua, Fakhreddine Karray, and Mohamed S. Kamel. 2017. “Recent Trends in Driver Safety Monitoring Systems: State of the Art and Challenges.” *IEEE Transactions on Vehicular Technology* 66 (6): 4550–63. https://doi.org/10.1109/TVT.2016.2631604.

Kuru, Kaya. 2021. “Conceptualisation of Human-on-the-Loop Haptic Teleoperation With Fully Autonomous Self-Driving Vehicles in the Urban Environment.” *IEEE Open Journal of Intelligent Transportation Systems* 2 (December):448–69. https://doi.org/10.1109/OJITS.2021.3132725.

Ledesma, Rubén Daniel, Silvana Andrea Montes, Fernando Martín Poó, and María Fernanda López-Ramón. 2015. “Measuring Individual Differences in Driver Inattention: Further Validation of the Attention-Related Driving Errors Scale.” *Human Factors* 57 (2): 193–207. https://doi.org/10.1177/0018720814546530/ASSET/IMAGES/LARGE/10.1177_0018720814546530-FIG2.JPEG.

Manchon, J. B., Mercedes Bueno, and Jordan Navarro. 2022. “How the Initial Level of Trust in Automated Driving Impacts Drivers’ Behaviour and Early Trust Construction.” *Transportation Research Part F: Traffic Psychology and Behaviour* 86 (April):281–95. https://doi.org/10.1016/j.trf.2022.02.006.

Morales-Alvarez, Walter, Oscar Sipele, Régis Léberon, Hadj Hamma Tadjine, and Cristina Olaverri-Monreal. 2020. “Automated Driving: A Literature Review of the Take over Request in Conditional Automation.” *Electronics (Switzerland)* 9 (12): 1–34. https://doi.org/10.3390/electronics9122087.

National Highway Traffic Safety Administration. 2017. “Automated Driving System 2.0: A Vision for Safety.” *U.S. Department of Transportation*. https://www.nhtsa.gov/vehicle-manufacturers/automated-driving-systems.

Payre, William, Jaume Perelló-March, and Stewart Birrell. 2023. “Under Pressure: Effect of a Ransomware and a Screen Failure on Trust and Driving Performance in an Automated Car Simulation.” *Frontiers in Psychology* 14 (March):1078723. https://doi.org/10.3389/FPSYG.2023.1078723/BIBTEX.

Petermeijer, Sebastiaan, Pavlo Bazilinskyy, Klaus Bengler, and Joost de Winter. 2017. “Take-over Again: Investigating Multimodal and Directional TORs to Get the Driver Back into the Loop.” *Applied Ergonomics* 62:204–15. https://doi.org/10.1016/j.apergo.2017.02.023.

Ramos, Marilia A., Christoph A. Thieme, Ingrid B. Utne, and A. Mosleh. 2020a. “Human-System Concurrent Task Analysis for Maritime Autonomous Surface Ship Operation and Safety.” *Reliability Engineering and System Safety* 195 (March):106697. https://doi.org/10.1016/j.ress.2019.106697.

Ramos, Marilia A., Christoph A. Thieme, Ingrid B. Utne, and Ali Mosleh. 2020b. “A Generic Approach to Analysing Failures in Human – System Interaction in Autonomy.” *Safety Science* 129 (September):104808. https://doi.org/10.1016/j.ssci.2020.104808.

Reason, James, Antony Manstead, Stradling Stephen, James Baxter, and Karen Campbell. 1990. “Errors and Violations on the Roads: A Real Distinction?” *Ergonomics* 33 (10–11): 1315–32. https://doi.org/10.1080/00140139008925335.

Reiman, Teemu, and Elina Pietikäinen. 2012. “Leading Indicators of System Safety – Monitoring and Driving the Organizational Safety Potential.” *Safety Science* 50 (10): 1993–2000. https://doi.org/10.1016/J.SSCI.2011.07.015.

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.

Sheehan, Cathy, Ross Donohue, Tracey Shea, Brian Cooper, and Helen De Cieri. 2016. “Leading and Lagging Indicators of Occupational Health and Safety: The Moderating Role of Safety Leadership.” *Accident Analysis and Prevention* 92 (July):130–38. https://doi.org/10.1016/j.aap.2016.03.018.

Shetty, Akhil, Hamidreza Tavafoghi, Alex Kurzhanskiy, Kameshwar Poolla, and Pravin Varaiya. 2022. “Automated Vehicle Safety and Deployment: Lessons from Human Crashes.” In *2022 International Conference on Connected Vehicle and Expo (ICCVE)*, 1–6. IEEE. https://doi.org/10.1109/ICCVE52871.2022.9742994.

Talebian, Ahmadreza, and Sabyasachee Mishra. 2022. “Unfolding the State of the Adoption of Connected Autonomous Trucks by the Commercial Fleet Owner Industry.” *Transportation Research Part E: Logistics and Transportation Review* 158 (February):102616.

Tener, Felix, and Joel Lanir. 2022. “Driving from a Distance: Challenges and Guidelines for Autonomous Vehicle Teleoperation Interfaces.” In *Conference on Human Factors in Computing Systems - Proceedings*, 1–13. New York, NY, USA: ACM. https://doi.org/10.1145/3491102.3501827.

Thorn, Eric, Shawn Kimmel, and Michelle Chaka. 2018. “A Framework for Automated Driving System Testable Cases and Scenarios.” *DOT HS 812 623*. www.ntis.gov.

Torc. 2021. “Torc Safety Report.” https://torc.ai/safety/vssa/.

TuSimple. 2021. “TuSimple Safety Report.” https://s202.q4cdn.com/364265561/files/doc_downloads/esg/TuSimple-Safety-Report.pdf.

Underwriters Laboratories. 2023. “ANSI/UL4600 - Evaluation of Autonomous Products.” Northbrook, IL.

United States. Department of Transportation. National Highway Traffic Safety Administration. National Center for Statistics and Analysis. 2023. "Traffic Safety Facts 2021 Data: Large Trucks," June. https://doi.org/10.21949/1503647.

Vanderhaegen, Frédéric. 2021. "Heuristic-Based Method for Conflict Discovery of Shared Control between Humans and Autonomous Systems - A Driving Automation Case Study." *Robotics and Autonomous Systems* 146 (December):103867. https://doi.org/10.1016/j.robot.2021.103867.

Wang, Chen, Yuanchang Xie, Helai Huang, and Pan Liu. 2021. "A Review of Surrogate Safety Measures and Their Applications in Connected and Automated Vehicles Safety Modeling." *Accident Analysis & Prevention* 157 (May): 106157. https://doi.org/10.1016/j.aap.2021.106157.

Warg, Fredrik, Martin Skoglund, and Matthew Sassman. 2021. "Human Interaction Safety Analysis Method for Agreements with Connected Automated Vehicles." In *2021 IEEE 94th Vehicular Technology Conference (VTC2021-Fall)*, 2021-Septe:01–07. IEEE. https://doi.org/10.1109/VTC2021-Fall52928.2021.9625202.

Winter, J. C.F. De. 2013. "Predicting Self-Reported Violations among Novice License Drivers Using Pre-License Simulator Measures." *Accident Analysis & Prevention* 52 (March):71–79. https://doi.org/10.1016/J.AAP.2012.12.018.

Wishart, Jeffrey, Steven Como, Maria Elli, Brendan Russo, Jack Weast, Niraj Altekar, and Emmanuel James. 2020. "Driving Safety Performance Assessment Metrics for ADS-Equipped Vehicles." *SAE International Journal of Advances and Current Practices in Mobility* 2 (5): 2881–99. https://doi.org/10.4271/2020-01-1206.

Xu, Runsheng, Yi Guo, Xu Han, Xin Xia, Hao Xiang, and Jiaqi Ma. 2021. "OpenCDA: An Open Cooperative Driving Automation Framework Integrated with Co-Simulation." *IEEE Conference on Intelligent Transportation Systems, Proceedings, ITSC* 2021-Septe:1155–62. https://doi.org/10.1109/ITSC48978.2021.9564825.

Xu, Runsheng, Hao Xiang, Xu Han, Xin Xia, Zonglin Meng, Chia-Ju Chen, Camila Correa-Jullian, and Jiaqi Ma. 2023. "The OpenCDA Open-Source Ecosystem for Cooperative Driving Automation Research." *IEEE Transactions on Intelligent Vehicles* 8 (4): 2698–2711. https://doi.org/10.1109/TIV.2023.3244948.

Yang, X. Jessie, Anuj Pradhan, Dawn Tilbury, and Lionel +"Jr" Robert. 2018. "Human Autonomous Vehicles Interactions: An Interdisciplinary Approach," April. http://deepblue.lib.umich.edu/handle/2027.42/142722.

Zhai, Siming, Shan Gao, Lin Wang, and Peng Liu. 2023. "When Both Human and Machine Drivers Make Mistakes: Whom to Blame?" *Transportation Research Part A: Policy and Practice* 170 (April):103637. https://linkinghub.elsevier.com/retrieve/pii/S0965856423000575.

Zhang, Bo, Joost de Winter, Silvia Varotto, Riender Happee, and Marieke Martens. 2019. "Determinants of Take-over Time from Automated Driving: A Meta-Analysis of 129 Studies." *Transportation Research Part F: Traffic Psychology and Behaviour* 64:285–307. https://doi.org/10.1016/j.trf.2019.04.020.

Zhang, Tao. 2020. “Toward Automated Vehicle Teleoperation: Vision, Opportunities, and Challenges.” *IEEE Internet of Things Journal* 7 (12): 11347–54. https://doi.org/10.1109/JIOT.2020.3028766.