Deriving human-system safety metrics for heavy-duty automated vehicle applications through Concurrent Task Analysis

Anna Cosmin-Spanoche[a,b*], Camila Correa-Jullian[a,b], Xin Xia[c], Ali Mosleh[b] , Jiaqi Ma[c],
[a]Department of Mechanical and Aerospace Engineering, University of California, Los Angeles, USA
[b]B. John Garrick Institute for the Risk Sciences, University of California, Los Angeles, USA
[c]Mobility Lab, University of California, Los Angeles, USA

Abstract: The use of automated driving technology in heavy-duty vehicles for commercial freight operations aims to increase efficiency and operational hours, as well as reduce traffic incidents. While there are currently over twenty companies actively developing HD-AV systems in the United States, the regulatory framework needed to implement these systems at a commercial level must address the unique safety risks that Automated Driving System (ADS) technology introduces. Potential HD-AV operations envision a team of human and machine agents, including the ADS, an onboard safety driver, a fleet operations centre, and, in some cases, an onboard safety operator. The complex interactions between these human and machine agents must be addressed when determining the system's safety requirements and design. Safety metrics usually focus on ADS performance, but to adequately inform system design and safety requirements, these metrics must also focus on human-system interactions. The data gleaned from these metrics can lead to improvements in the human-machine interface (HMI), warnings and alerts, and task allocation between the safety drivers and ADS. This work proposes an approach to derive human-system interaction safety metrics based on Concurrent Task Analysis (CoTA), a method built to study human and automated system interactions from a task decomposition and success-oriented analysis perspective. The CoTA method, based on the Information, Decision and Action (IDA) cognitive model, is used to model the tasks performed by different agents and study the interactions between them. In turn, this can inform procedure development, identify contributing task errors and propagation mechanisms, and identify the critical tasks needed for success of a system. This work uses CoTA to identify the most critical tasks of safety drivers in an HD-AV operation, and with this information discusses safety-related human-system interaction metrics to inform HD-AV system development.

Keywords: automated driving systems, human-system interactions, safety metrics, concurrent task analysis

1. INTRODUCTION

Automated driving technology has received significant attention in a variety of transportation-related applications, including passenger transport, onboard driver assistance features, and commercial applications. Automated driving systems (ADS) and related technologies may have numerous transportation, commercial, and legal impacts, and the degree to which they are adopted in the future resides heavily on in-depth identification and response to the risks associated with them. Automation is categorized by the Society of Automotive Engineers into six levels, ranging from 0-5. These distinctions are based on the degree to which human and autonomous agents perform the Dynamic Driving Task (DDT). For Levels 0-2, the human driver performs the DDTs supported by driving assistance features. For Levels 3-4, the ADS performs the DDT limited to a specific Operational Design Domain (ODD) with different levels of fallback involvement from a human agent. At Level 5, the ADS is not restricted to a specific ODD and is expected to perform all DDTs independently [1].

The incorporation of ADS technology into heavy-duty commercial transport operations is motivated by a desire to reduce traffic collisions and incidents related to human errors, as well as increasing efficiency and operational hours [2]. Heavy-duty vehicles are defined as vehicles weighing over 26,001 lbs. and can take the form of buses, construction vehicles, or trucks, among other uses. The most common use of heavy-duty vehicles is in commercial operations using trucks, which are estimated to carry 70% of yearly freight tonnage in the United States [3]. Various reports have discussed the impact of replacing human drivers with automated technology, and the estimated rate of crash reduction by adopting this technology ranges from 50% [4] to 90% [5]. Additionally, since the operational hours of commercial trucking operations are limited by the Hours-of-Service regulation of the US Department of Transportation (US DOT), increasing autonomy can potentially extend those hours towards eventual 24/7 operation. By also incorporating platooning and live traffic data, fuel efficiency could be improved, leading to decreased operational costs [3]. HD-AV operations can potentially cover a range of trucking applications, including middle-mile, drayage, and long-haul, which each present unique challenges, tasks, operational profiles, and safety requirements. While there are over twenty companies in various stages of HD-AV development internationally, this technology has not reached deployment on public roads yet. Therefore, efforts must be directed towards understanding and assessing the new risks that arise by increasing the number of heavy-duty vehicles on the road, expanding their operational hours, and the use of emerging technologies.

HD-AV operations currently envision operations within the range of Levels 2-4 of driving automation, and many planned operations include human agents in some capacity [2]. These human agents can take the roles of remote operators, safety drivers, or safety operators. The role of these human agents can incorporate performing sections of driving prior to entering the ODD, conducting control transitions, and responding to emergency scenarios. Therefore, it is critical to assess the safety implications of these human-ADS interactions. To quantify the safety of HD-AV systems, a set of metrics including the unique aspects of human-system interaction must be defined. Guidance for the creation and implementation of ADS safety and performance metrics are documented in standards such as UL4600, ISO26262, and SOTIF [6]. However, these do not fully consider the role of safety drivers in the current planned use cases of HD-AVs. While current discussions about safety drivers in HD-AVs are limited to testing phases prior to public deployment, safety drivers are likely to remain present during nominal operations due to legal and regulatory requirements. Metrics to assess ADS performance are mostly limited to assessing the functional safety of the. Traditional safety metrics, referred to as *lagging* metrics, record the occurrence of failure events such as accidents, injuries, and fatalities. Examples of these metrics include incident rate per miles driven or number of fatalities. However, there has been an increase of research into constructing *leading* metrics, which collect data related to safety-relevant events that do not lead to a catastrophic event. These surrogate safety metrics (SSMs) include measures like Time-To-Collision (TTC), Deceleration Rate to Avoid the Crash (DRAC), and DeltaV, which aim to quantify interactions with other road vehicles during ‘near misses’ and non-failure events [7]. Leading metrics may provide further indications of system safety, allowing for standardization and comparison of non-failure events before the vehicles enter public deployment and high-severity incidents occur [8].

A systematic model-based approach is needed to construct metrics related to human-system interactions in HD-AV operations. Qualitative model-based risk assessment methods such as CoTA and System-Theoretic Process Analysis (STPA) can serve as a basis to conduct hazard identification for quantitative risk models used in probabilistic risk assessment (PRA) such as ESDs and Fault Trees (FTs). The human-system interaction in autonomy (H-SIA) method using CoTA models was initially applied to analyse autonomous maritime vessels with remote operators [9]. This work presents a methodology for deriving human-system interaction safety metrics based on the CoTA model, and it is applied to a case study specifically for HD-AVs. The HD-AV system is modelled into agents and operational scenarios, and ESDs and CoTAs are used to construct metrics highlighting human-system interaction.

2. SAFETY METRIC DERIVATION METHODOLOGY

An operational safety hazard identification methodology developed for complex socio-technical systems is employed to develop metrics for human-system interaction. This structured process combines several risk assessment techniques, including ESDs, CoTA, FTs, and STPA to analyse the interactions between agents in a complex system and identify hazards. It was 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. This hazard identification method is divided into three stages: system modelling, scenario modelling, and hazard identification [10]. The present work takes advantage of the first stages of the hazard identification methodology for the HD-AV scenario, applying the ESD and CoTA steps to derive metrics specifically for safety driver-ADS interactions. Stage 1 includes modelling the system by defining agents and their high-level tasks (Step 1) and defining operational phases and the transitions between them (Step 2). Stage 2 (scenario modelling) involves documenting operational phases through ESDs (Step 3) and modelling tasks and interactions using CoTA (Step 4). Stage 1 was implemented for the HD-AV system, and a list of preliminary human-interaction metrics based on the functional breakdown and operational phases were developed [11]. In this work, Steps 3 and 4 from Stage 2 are implemented to identify and refine the metrics employing a structured, task-oriented approach. Further details on the method and extended results can be found in [10].

2.1. System Modelling

The first stage of the method involves modelling the system by performing a breakdown for each agent and defining operational phases. A generic model of HD-AV operations was developed by analysing a current sample of companies developing and testing in the HD-AV space and following NHTSA guidance for ADS system design [12]. The companies observed for reference fleet creation include Aurora, Kodiak Robotics, and Torc Robotics [13-15]. The relevant agents in the reference HD-AV are the ADS, safety driver, safety operator (optionally), and fleet operations centre (FOC).

The ADS involves the hardware and software that performs the DDT when engaged. The DDT includes planning and executing latitudinal and longitudinal vehicle movements (e.g. steering, braking, and lane changes). The ADS operates nominally at a Level 4 of driving automation; namely, it can implement fallback procedures without a need for take-over by the safety driver while it is within its ODD. Its ODD is limited by geofenced maps and is restricted to public highway roads during clear visibility and fair-weather conditions. As a result, due to rapidly developing weather events, geofencing, or mapping errors, ODD breaches can occur during operation, requiring timely intervention from the safety driver to ensure safety. The safety driver is a trained commercial vehicle operator who sits in the driver's seat, performs the DDT for sections of the ride, and is responsible for disengaging the ADS and taking control of the vehicle in the case of an ODD exit or other emergency scenario. The safety driver receives data from the ADS about vehicle status and any potential malfunctions through the human-machine interface (HMI) and has communication pipelines with the FOC. The FOC is a location in which hired operators monitor the HD-AV fleet in a control room environment. Each remote operator may be tasked to monitor multiple HD-AVs through a dashboard and provide warnings to the onboard safety drivers. In addition, remote operators receive incident and ODD breach notifications automatically and play a role in incident response. It is possible that HD-AV operations also contemplate on-board safety operators or a trained individual in the passenger's seat that can assist the safety driver in monitoring operations. A brief description of each agent and corresponding high-level tasks is provided in Table 1. Although present in some reference fleet operations, the safety operator is an optional agent and has not been included in further analysis.

Table 1: Agents in HD-AV System.

Agent Name	Description	Selected High Level Tasks
ADS	The software and hardware responsible for performing the DDT within the limits of the ODD, nominal Level 4 autonomy.	Performing the DDT while activated, monitoring the safety driver, requesting control transitions when needed.
Safety Driver (DRI)	Commercial vehicle operator with valid Commercial Driver's License (CDL) who has completed training for commercial driving and ADS operations.	Driving the vehicle outside its ODD, engaging and disengaging the automated driving phase.
Safety Operator	Additional onboard human agent onboard in the passenger's seat.	Monitoring road conditions and the state of the HD-AV, communicating with safety driver.
FOC	Physical location where operators monitor the HD-AV fleet in a control room environment.	Backup support for vehicle and driver monitoring, assist with incident response.

The generic HD-AV operation is decomposed into distinct phases that occur during a normal shift. The identified phases are Pre-Shift Inspection, Manual Driving, Manual Driving—Ready to Engage, Automated Driving Engaged, and Fallback/MRC. In the reference fleet, the ADS can only be engaged within its ODD, encompassing "middle mile" operations on highway roads. Hence, an operational shift begins with the vehicle and ADS being inspected and approved (Pre-Shift Inspection) and continues with the safety driver manually driving until reaching a highway on-ramp within the ODD (Manual Driving). When the vehicle enters the ODD and the ADS determines a handover is feasible, the state transitions to Ready to Engage. From here, the safety driver can decide to perform an ADS hand-over and engage automated driving (Automated Driving Engaged). At any point during the Automated Driving Engaged operational phase, the safety driver can perform a take-over to control of the DDT. It can be triggered by an ODD breach, perceived collision risk, or any other reason. Additionally, during all operational phases, elements such as vehicle diagnostics, ADS status, and road conditions are monitored by the ADS, safety driver, and FOC – with varying degrees of details. If these is a safety-related issue or traffic incident, a DDT fallback is triggered and the ADS is prompted to enter MRC, or the safety driver can perform a take-over (Fallback/MRC). The FOC is also notified in this case for potential incident response. The transitions between Manual Driving and Automated Driving Engaged (ADS control transitions), and the Fallback/MRC stage are analysed through ESDs and CoTA. More details including a diagram of the operational breakdown can be found in [11].

2.2 Scenario Modelling through ESD

The third step of the hazard identification methodology is developing ESDs representing each operational phase and the transitions. Here the primary focus is developing qualitative ESDs representing the transitions between Manual Driving, Ready to Engage, and Automated Driving operational phases. The critical scenarios modelled were the safety driver requesting an ADS hand-over (ESD 1.1), the safety driver performing a take-over (ESD 1.2), the manual driving phase transitioning to the ready to engage phase (ESD 2.1), and the ADS vehicle approaching the ODD limits (ESD 2.2). ESDs 2.1–2.2 contain transition states to ESDs 1.1–1.2 depending if a take-over or hand-over is requested. The end states in the ESDs denote a successful trip (ES1), a delayed trip (EF2), or a collision risk (EF3). With real-world and simulation data, quantitative probabilities and risk levels could be attached to the events and end states in the ESD. If an end state serves as an initiating event for a different ESD, these are noted as transition states. For instance, in ESD 2.2, the initiating event considers the ADS operating within the ODD as shown in Fig. 1.

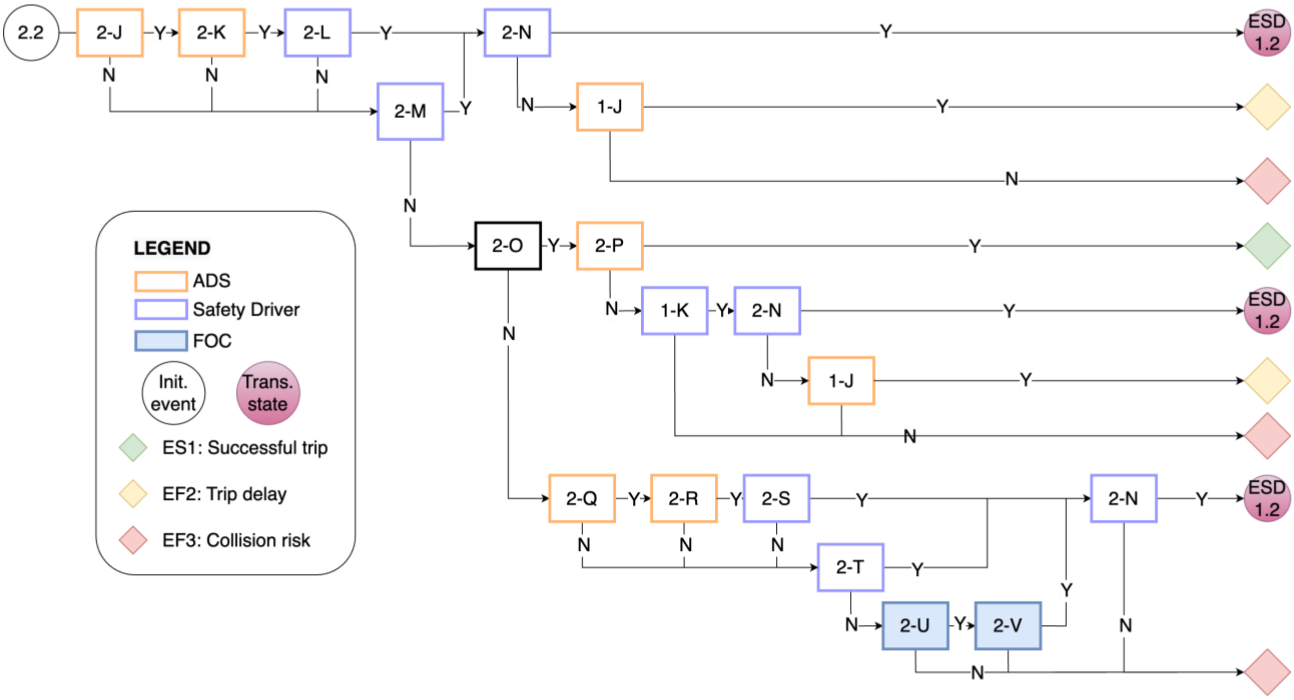

Figure 1: ESD 2.2 - Vehicle approaches limits of ODD.

The event sequence pictured in Fig. 1 and described in Table 2 initiates with 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 (FMS), 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 FMS. 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 take-over. If the FOC does not detect the ODD breach, there is an issue in performing the take-over, or FMS fails, this may lead to an unmitigated risk of collision.

Table 2: 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
2-M	Driver detects ODD limit approach.	DRI
2-N	Driver performs takeover of vehicle.	DRI
1-J	Vehicle can implement FMS.	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 vehicle has exited 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 driver that vehicle has exited ODD.	FOC
ESD 1.2	Driver performs take-over.	DRI

2.3 Identifying critical tasks through CoTA

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 analyse 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 [16]. These tasks are categorized based on the IDA (Information, Decision, Action) cognitive model [17]. 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), but it has been also applied to the maritime industry in the context of Autonomous Ships (AS). Each task and subtask in CoTA is categorized into a series of plans which denote the order in which the tasks are carried out sequentially (e.g., 1->2), in parallel (e.g., 1//2), triggered by other tasks (e.g., 1—2) or performed exclusively (e.g., 1 or 2).

CoTA models were developed for the Safety Driver, ADS, and FOC agents in this system. For illustration purposes, only a simplified Safety Driver CoTA is presented in Fig. 2, but as needed, connections to other agent CoTAs are shown. The role of the safety driver in the HD-AV was decomposed into six high-level tasks described in Table 3. 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 receives transition requests from the ADS, and 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 hand-over requests, and transmits and receives driver alerts.

Table 3. High-Level Safety Driver 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 OEDR functions, following local traffic rules, and, if necessary, implementing tactical manoeuvres.
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.
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, here the driver manages the vehicle control transitions. The driver may request control transitions, i.e., driver-initiated handovers or take-overs and respond to system-initiated requests.
6	Communicate with FOC	Parallel	At all operational phases, the safety driver communicates with the fleet operations centre. For this task, the safety driver receives communications from the FOC, plans a response, and then responds to the FOC.

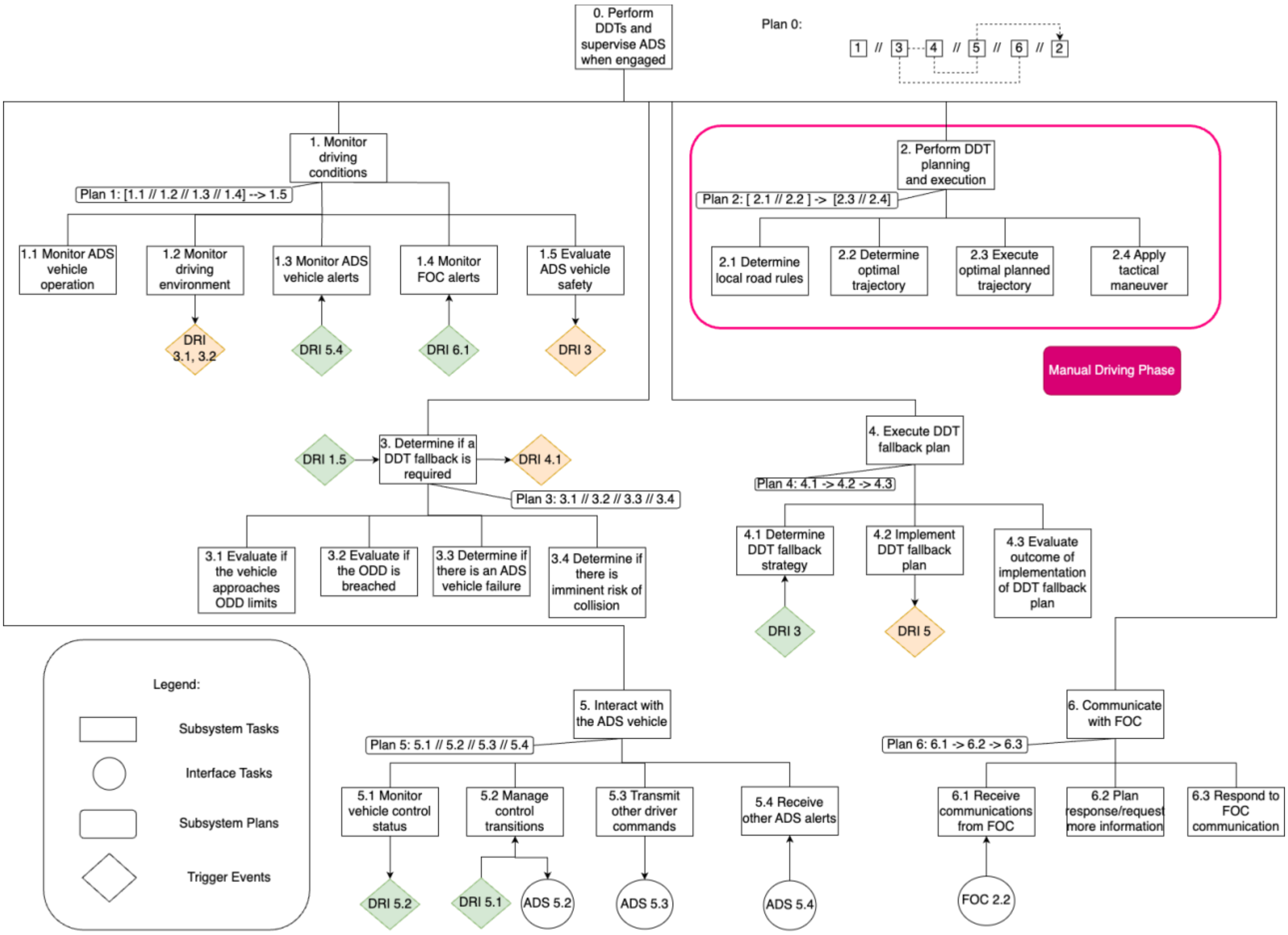
0. Perform DDTs and supervise ADS when engaged
Plan 0:
1 // 3 4 // 5 // 6 // 2
1. Monitor driving conditions
Plan 1: [1.1 // 1.2 // 1.3 // 1.4] --> 1.5
1.1 Monitor ADS vehicle operation
1.2 Monitor driving environment
1.3 Monitor ADS vehicle alerts
1.4 Monitor FOC alerts
1.5 Evaluate ADS vehicle safety
DRI 3.1, 3.2
DRI 5.4
DRI 6.1
DRI 3
2. Perform DDT planning and execution
Plan 2: [2.1 // 2.2] -> [2.3 // 2.4]
2.1 Determine local road rules
2.2 Determine optimal trajectory
2.3 Execute optimal planned trajectory
2.4 Apply tactical maneuver
Manual Driving Phase
DRI 1.5
3. Determine if a DDT fallback is required
DRI 4.1
Plan 3: 3.1 // 3.2 // 3.3 // 3.4
3.1 Evaluate if the vehicle approaches ODD limits
3.2 Evaluate if the ODD is breached
3.3 Determine if there is an ADS vehicle failure
3.4 Determine if there is imminent risk of collision
4. Execute DDT fallback plan
Plan 4: 4.1 -> 4.2 -> 4.3
4.1 Determine DDT fallback strategy
4.2 Implement DDT fallback plan
4.3 Evaluate outcome of implementation of DDT fallback plan
DRI 3
DRI 5
Legend:
Subsystem Tasks
Interface Tasks
Subsystem Plans
Trigger Events
5. Interact with the ADS vehicle
Plan 5: 5.1 // 5.2 // 5.3 // 5.4
5.1 Monitor vehicle control status
5.2 Manage control transitions
5.3 Transmit other driver commands
5.4 Receive other ADS alerts
DRI 5.2
DRI 5.1
ADS 5.2
ADS 5.3
ADS 5.4
6. Communicate with FOC
Plan 6: 6.1 -> 6.2 -> 6.3
6.1 Receive communications from FOC
6.2 Plan response/request more information
6.3 Respond to FOC communication
FOC 2.2

Figure 2: Simplified CoTA – Safety Driver.

As shown in Tables 3 and 4, 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 take-over request or emergency stop request, but the ADS fallback leads to a Minimal Risk Condition (MRC) or Stable Stopped Condition (SSC), which are initiated by the vehicle and can lead to an operational delay. The interactions between the ADS and safety driver in the driver-initiated control transition task are demonstrated in the CoTAs in Fig. 3 and 4, which show the driver tasks that lead to ADS tasks and vice versa (see Table 5). For instance, for the "Manage driver-initiated takeovers" subtask, the driver first determines if a takeover is needed and then performs the takeover. This leads to the ADS task of detecting the driver take-over input, which then leads to the task of determining if the driver is in control of the vehicle. The CoTA provides a structured way to observe agent interactions on the same hierarchy and assess which of these interactions can point to operational risks.

Table 4. High-Level ADS 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 if 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 by 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 MRC or SSC. 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.

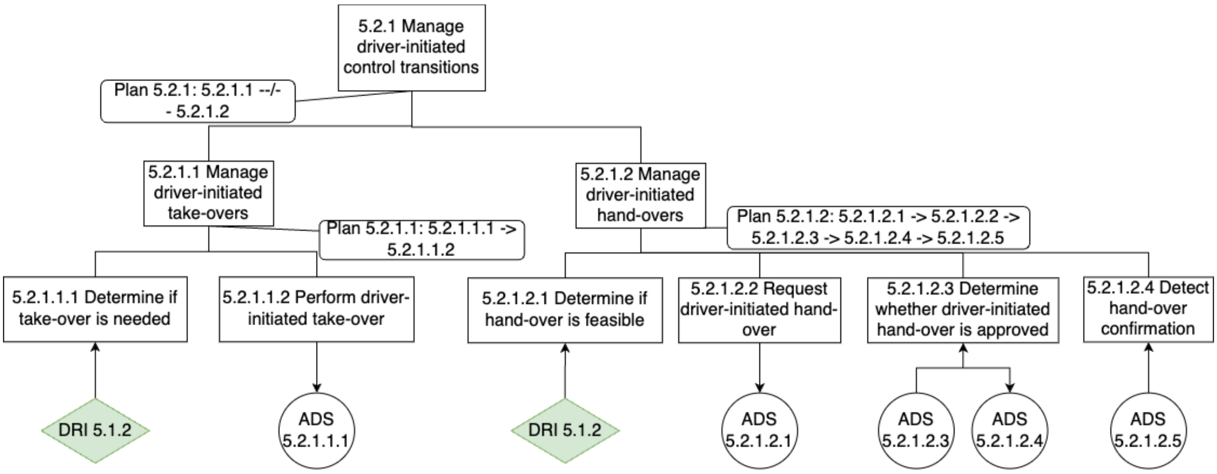
5.2.1 Manage driver-initiated control transitions
Plan 5.2.1: 5.2.1.1 --/- - 5.2.1.2
5.2.1.1 Manage driver-initiated take-overs
Plan 5.2.1.1: 5.2.1.1.1 -> 5.2.1.1.2
5.2.1.2 Manage driver-initiated hand-overs
Plan 5.2.1.2: 5.2.1.2.1 -> 5.2.1.2.2 -> 5.2.1.2.3 -> 5.2.1.2.4 -> 5.2.1.2.5
5.2.1.1.1 Determine if take-over is needed
5.2.1.1.2 Perform driver-initiated take-over
5.2.1.2.1 Determine if hand-over is feasible
5.2.1.2.2 Request driver-initiated hand-over
5.2.1.2.3 Determine whether driver-initiated hand-over is approved
5.2.1.2.4 Detect hand-over confirmation
DRI 5.1.2
ADS 5.2.1.1.1
DRI 5.1.2
ADS 5.2.1.2.1
ADS 5.2.1.2.3
ADS 5.2.1.2.4
ADS 5.2.1.2.5

Figure 3: DRI Task 5.2.1 CoTA decomposition.

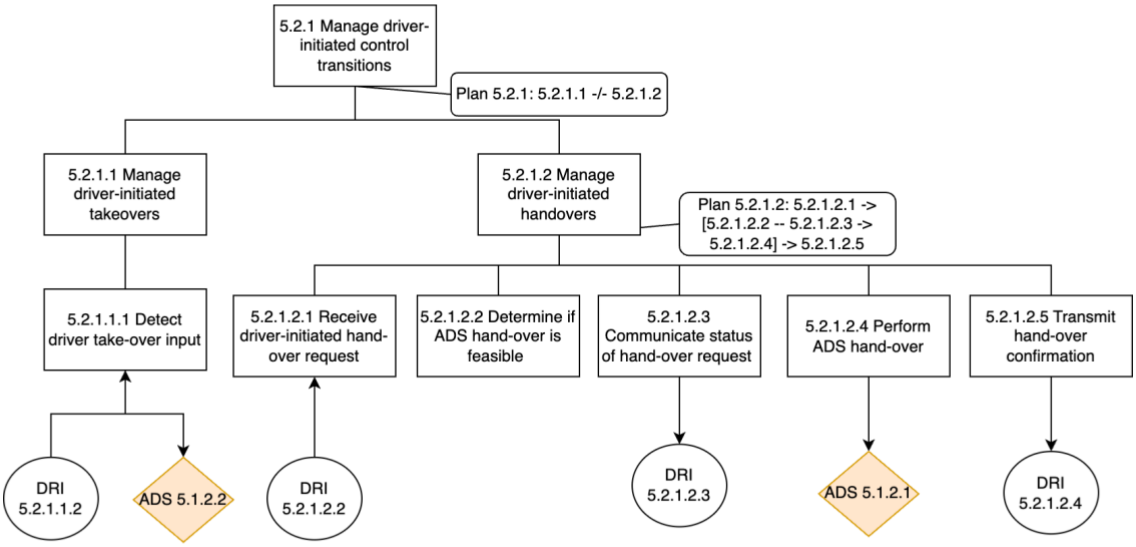
5.2.1 Manage driver-initiated control transitions
Plan 5.2.1: 5.2.1.1 -/- 5.2.1.2
5.2.1.1 Manage driver-initiated takeovers
5.2.1.2 Manage driver-initiated handovers
Plan 5.2.1.2: 5.2.1.2.1 -> [5.2.1.2.2 -- 5.2.1.2.3 -> 5.2.1.2.4] -> 5.2.1.2.5
5.2.1.1.1 Detect driver take-over input
5.2.1.2.1 Receive driver-initiated hand-over request
5.2.1.2.2 Determine if ADS hand-over is feasible
5.2.1.2.3 Communicate status of hand-over request
5.2.1.2.4 Perform ADS hand-over
5.2.1.2.5 Transmit hand-over confirmation
DRI 5.2.1.1.2
ADS 5.1.2.2
DRI 5.2.1.2.2
DRI 5.2.1.2.3
ADS 5.1.2.1
DRI 5.2.1.2.4

Figure 4: ADS Task 5.2.1 CoTA decomposition.

Table 5: 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

2.4 Derived Safety Metrics

In observing failure paths from the ESD and the tasks that leading 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. A selection of the proposed metrics for the modelled HD-AV system measuring control transitions and alerts are summarized in Tables 6 and 7. Ideally, these metrics can be tracked in HD-AV simulation and testing to inform design of components like HMI and operational tasks in initial stages and validate the system in later testing stages. Most of these metrics would be collected from the ADS data log, which keeps track of sensor and alarm data, operational phase changes including take-over and hand-over 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 certain decisions made by the safety drivers and FOC operators assess the quality of the ADS post-incident response.

Table 6: Control Transition Metrics.

#	Name	Definition	Unit	Data source	Origin
1	Rate of Successful Driver-Initiated Handovers	Ratio of successful driver handovers to total number of handover requests.	%	Data log	DRI CoTA 5.2.1.2.4, ADS CoTA 5.2.1.2.5, ESD 1.1
2	Rate of Successful Driver-Initiated Take-overs	Ratio of successful driver take-overs to total number of take-over attempts.	%	Data log	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	DRI CoTA 5.2.2.2.3, ADS CoTA 5.2.2.2.5
4	Rate of Successful System-Initiated Take-overs	Ratio of successful driver take-overs to total number of system-initiated take-over requests.	%	Data log	DRI CoTA 5.2.2.1.2, ADS CoTA 5.2.2.1.3
5	Rate of ADS Hand-over Approval	Ratio of driver approval of system-initiated handovers to system-initiated hand-over requests.	%	Data log	DRI CoTA 5.2.2.2.2
6	Reason for Driver-Initiated Take-over	Category for reason safety driver initiated a take-over (e.g. lack of trust, unnoticed ODD breach).	n/a	Survey or interview	ESD 1.2, DRI CoTA 5.2.1.1
7	Reason for System-Initiated Take-over	Category for reason ADS initiated a take-over request (e.g. ODD breach, collision risk).	n/a	Data log	ADS CoTA 5.2.2.1.1
8	Quality of Take-over (TTC-Based)	Minimum time to collision and maximum resulting lateral and longitudinal acceleration after the initiated take-over request.	sec	Data log	Literature/ system model
9	Quality of Take-over (Dynamics-Based)	Maximum resulting weighted sum of lateral and longitudinal acceleration after the initiated take-over request.	m/s^2	Data log	Literature/ system model
10	Quality of Take-over (TOT-Based)	Take-over time (TOT) interval between take-over request (TOR) and the driver’s first manoeuvre	sec	Data log	Literature/ system model

For instance, the metric “Rate of Successful Driver-Initiated Take-overs” was determined by observing the interaction between DRI and ADS Task 5.2.1.1. A failure in detection of driver take-over input would lead to an unsuccessful take-over, 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. Similarly, alert-related metrics were used to measure the ratio of alerts arising from diverse sources (i.e. vehicle-related malfunctions, driver monitoring system, ODD breaches, etc.) as well as the ratio of alerts not responded to by the safety driver. Metrics were also developed for the following groups: Incident, Fallback, and Human-ADS Trust. The Incident metrics track incident rate per miles driven, and these incident 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. The Fallback metrics refer to rates of FMS, emergency stops, and cases of ADS fallback resulting from driver inaction. The Human-ADS Trust metrics consist of measurements of disagreement with ADS manoeuvres 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 [18].

Table 7: Alert-Related Metrics.

#	Name	Definition	Unit	Data source	Origin
6	Alerts Resulting from Vehicle-Related Malfunctions	Ratio of alerts coming from vehicle sensor, ADS or vehicle malfunction to total number of alerts	%	Data log	DRI CoTA 5.4.2
7	Alerts Resulting from Onboard Safety Driver	Ratio of alerts generated by the driver monitoring system to total number of alerts	%	Data log	DRI CoTA 5.4.3
8	Alerts Resulting from Enviroment	Ratio of alerts coming from road conditions and ODD breach to total number of alerts	%	Data log	DRI CoTA 5.4.1
9	Alerts Not Acted On	Ratio of alerts not responded to by safety driver to total number of alerts	%	Data log	ESD 5.4

3. 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 being at HMI display, alert design, or driver training level. 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 [19]. 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.

4. CONCLUSION

With the increased interest in incorporating ADS into heavy-duty commercial operations, the role that human-autonomy teams play has not been fully assessed in the HD-AV framework. Due to regulatory and legal framework, safety drivers will likely continue to be involved in HD-AV operations 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 HMI in the vehicle and inform design of operational tasks. Additionally, metrics can assess trends or point to needed changes during road testing stages and eventual public deployment. This work demonstrates a methodology to use 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 STPA and other methods to provide alternative characterization of the HD-AV systems, leading towards a more comprehensive hazard identification analysis and surrogate safety metric construction. The metric derivation methodology 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/miles driven and used to make decisions.

Acknowledgements

This research is partly supported by the University of California Institute of Transportation Studies. The work presented in this paper remains the sole responsibility of the authors.

References

[1] SAE International, ‘Taxonomy and Definitions for Terms Related to Driving Automation Systems for On-Road Motor Vehicles’, *SAE Standard J3016 APR2021*, 2021.

[2] A. K. Bhoopalam, R. van den Berg, N. Agatz, and C. G. Chorus, ‘The long road to automated trucking: Insights from driver focus groups’, *Transp Res Part C Emerg Technol*, vol. 156, Nov. 2023, doi: 10.1016/j.trc.2023.104351.

[3] A. Talebian and S. Mishra, ‘Unfolding the state of the adoption of connected autonomous trucks by the commercial fleet owner industry’, *Transp Res E Logist Transp Rev*, vol. 158, p. 102616, Feb. 2022, doi: 10.1016/J.TRE.2022.102616.

[4] A. Shetty, H. Tavafoghi, A. Kurzhanskiy, K. Poolla, and P. Varaiya, ‘Automated Vehicle Safety and Deployment: Lessons from Human Crashes’, in *2022 International Conference on Connected Vehicle and Expo (ICCVE)*, IEEE, Mar. 2022, pp. 1–6. doi: 10.1109/ICCVE52871.2022.9742994.

[5] J.-François. Bonnefon, *The car that knew too much: Can a machine be moral?* MIT Press, 2021.

[6] UL Standards and Engagement, ‘Standard for Safety for the Evaluation of Autonomous Products, UL 4600’, 2023.

[7] C. Wang, Y. Xie, H. Huang, and P. Liu, ‘A review of surrogate safety measures and their applications in connected and automated vehicles safety modeling’, *Accid Anal Prev*, vol. 157, Jul. 2021, doi: 10.1016/j.aap.2021.106157.

[8] T. Reiman and E. Pietikäinen, ‘Leading indicators of system safety - Monitoring and driving the organizational safety potential’, *Saf Sci*, vol. 50, no. 10, pp. 1993–2000, Dec. 2012, doi: 10.1016/j.ssci.2011.07.015.

[9] M. A. Ramos, C. A. Thieme, I. B. Utne, and A. Mosleh, ‘A generic approach to analysing failures in human – system interaction in autonomy’, *Saf Sci*, vol. 129, p. 104808, Sep. 2020, doi: 10.1016/J.SSCI.2020.104808.

[10] C. Correa-Jullian, M. Ramos, A. Mosleh, and J. Ma, ‘Operational safety hazard identification methodology for automated driving systems fleets’, *Proc Inst Mech Eng O J Risk Reliab*, 2024, doi: 10.1177/1748006X241233863.

[11] A. Cosmin-Spanoche, C. Correa-Jullian, X. Xia, A. Mosleh, and J. Ma, ‘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.

[12] National Highway Traffic Safety Administration, ‘Automated Driving Systems 2.0: A Vision for Safety’, 2017.

[13] Aurora, ‘Voluntary Safety Self-Assessment’, 2021.

[14] Kodiak, ‘Kodiak Safety Report 2020’, 2020.

[15] Torc, ‘Innovating Safety and Efficiency - Torc Safety Report’, 2021.

[16] M. A. Ramos, C. A. Thieme, I. B. Utne, and A. Mosleh, ‘Human-system concurrent task analysis for maritime autonomous surface ship operation and safety’, *Reliab Eng Syst Saf*, vol. 195, Mar. 2020, doi: 10.1016/j.ress.2019.106697.

[17] Y. H. J. Chang and A. Mosleh, ‘Cognitive modeling and dynamic probabilistic simulation of operating crew response to complex system accidents. Part 5: Dynamic probabilistic simulation of the IDAC model’, *Reliab Eng Syst Saf*, vol. 92, no. 8, pp. 1076–1101, Aug. 2007, doi: 10.1016/j.ress.2006.05.012.

[18] X. J. Yang, A. K. Pradhan, D. Tilbury, and L. Robert, ‘Human Autonomous Vehicles Interactions: An Interdisciplinary Approach’, 2018.

[19] R. Xu *et al.*, ‘The OpenCDA Open-Source Ecosystem for Cooperative Driving Automation Research’, *IEEE Transactions on Intelligent Vehicles*, vol. 8, no. 4, pp. 2698–2711, Apr. 2023, doi: 10.1109/TIV.2023.3244948.