Exploring Performance-Shaping Factors for Human-Autonomy Teams in Automated Driving System Operations

Camila Correa-Jullian[a*], Marilia Ramos[a], Ali Mosleh[a], Jiaqi Ma[b]

[a]The B. John Garrick Institute for the Risk Sciences, University of California Los Angeles, United States
[b]Mobility Lab, University of California Los Angeles, United States

Abstract: Autonomous systems and automated technologies are becoming increasingly prevalent across multiple industries and applications, such as aviation and air traffic control, marine transportation, and railway industries. In many of these complex systems, human operator teams frequently focus on monitoring and supervising the system's operation, acting as a safety barrier in emergencies. The need to study Human-Autonomy Team's (HAT's) performance in the case of automation failure and limited decision-making *explainability* may become more important as interactions between humans and machine agents diversify to non-expert systems. This is the case for drivers on board vehicles equipped with highly Automated Driving Systems (ADS). This work explores the applicability of Performance Shaping Factors (PSFs) used in Human Reliability Analysis (HRA) models to the HATs present in ADS operations. This work identifies potential factors influencing the performance of both human and automated agents in ADS operations to apply team performance models such as the Information, Decision and Action in Crew (IDAC) context. This work focuses on the relationship, tasks, and challenges drivers face when interacting with vehicles equipped with advanced ADS. It highlights the role that individual, system, team, and scenario-related factors play in the overall system's safety.

Keywords: human-autonomy teams, automated driving systems, performance shaping/influencing factors

1. INTRODUCTION

The term Human-Autonomy Teams (HATs) describes teams in which at least one member meets the definition of an *autonomous machine agent* and acts interdependently with other human members to achieve a collective goal [1]. As autonomous functions, capacities, and systems increase across diverse industrial and commercial developments, this hybrid team concept provides many opportunities to leverage decades of human teams' research and explore the emerging dynamics of HATs outside traditional industrial or control room operation contexts. This is the case of Automated Driving Systems (ADS), where human-system interactions have significant impacts on the overall system's safety. ADS technologies are expected to play a significant role in the transportation ecosystem, either for the commercial transport of goods and passengers, or for personal use. Along with the complex technical challenges this implies, it is the role of developers, manufacturers, and regulators to consider the role humans play in ADS safety as drivers, passengers, and fellow road users.

Decades of human-vehicle interaction research efforts in human factors and psychology have studied human performance under different driving conditions and external stimuli [2]. More recently, these studies have also investigated how human performance is altered by the presence of automated driving features [3]. As in the case of other autonomous systems, research trends suggest evolving from human-system interaction schemes towards human-vehicle collaboration, implying treating human-system agents as a team [4]. Currently, automated driving technology is organized into a six-level scale [5]. These levels are broadly divided into driver support features (Levels 0-2) – commercially referred to as Advanced Driving Assistance Systems (ADAS) – and Automated Driving Systems (Levels 3-5). This division of Dynamic Driving Tasks (DDTs) is based on the task allocation between the human and the automated driving technology. From Level 3 (L3) onwards, the DDTs are progressively transferred from the human driver to the ADS. However, at L3, the human driver is still expected to act as a *fallback-ready user*, meaning they are ultimately responsible for intervening in the vehicle's actions upon the request of the ADS, or preemptively to avoid emergency situations.

The control transitions between the driver and the ADS usually occur when approaching the exit of the Operational Design Domain (ODD) or in unexpected situations [6], giving the driver a short time budget to assess the situation and react accordingly to reach a safe state. At Level 4 (L4), the ADS are expected to perform DDT fallbacks and achieve a Minimal Risk Condition (MRC) autonomously; hence, the user is not expected to monitor or intervene in the vehicle's actions *while* the vehicle remains within the ODD. Currently, driverless L4 ADS vehicles have emerged as passenger mobility service providers, usually supported by a remote operation center with limited monitoring and incident management capacities. Even at high levels of automation, it is highly likely humans will continue interacting with these systems as drivers, operators, and/or fellow road users. As vehicles equipped with ADS capacities increase their share in the market and on the roads, it is crucial to advance the state of understanding of HATs in ADS contexts.

In complex engineering systems, Human Reliability Analysis (HRA) research provides valuable qualitative and quantitative insights into Risk Assessments to improve system, procedure, standard, and regulation design. To model human performance and to quantitively express Human Error Probabilities (HEPs), different HRA models rely on Performance Shaping Factors or Performance Influencing Factors (PSF/PIF) to describe factors affecting performance, including detailed cognitive process elements at individual and team levels, the effect of HSI design and other contextual elements [7]. For example, the cognitive model of Information, Decision, and Action in Crew context (IDAC) was developed to model Nuclear Power Plant (NPP) operator crews in control room environments [8]. IDAC was developed based on numerous relevant findings from cognitive psychology, behavioral science, neuroscience, human factors, social science, field observations, and various first- and second-generation HRA methodologies. In addition to a cognitive model that can be used in HRA methodologies, such as Phoenix [9], IDAC can also be implemented as an HRA model. IDAC models how an operator performs information processing (I), problem-solving and decision-making (D), and acts (A) within the context of a crew (C), while influenced by a set of internal and external PSFs. After its first publication, the IDAC model has been adapted and expanded into different versions and simulation frameworks, incorporating new knowledge, simulation tools, and updating PSF taxonomies to express additional contextual factors affecting team performance [9], [10]. Most PSF taxonomies have been developed in NPPs contexts but have also been extended to other industries and applications [11], [12].

HRA principles and elements, such as PSFs, can benefit the analysis of HAT and risk assessments of ADS operations. This work explores the applicability of IDAC-based individual, team, and scenario factors to driver-ADS teams in conditional driving automation contexts (L3). In this setting, one of the most critical tasks is successfully reacting to automation-initiated control transition triggers, commonly referred to as takeover requests. A review of selected PSFs potentially applicable to ADS driver-ADS team contexts is presented, followed by discussion of how future work and data collection initiatives may be pursued to derive risk assessment insight at both design and operation time.

2. MODELING HUMAN-AUTONOMY TEAMS IN AUTOMATED DRIVING SYSTEMS

In complex systems, the traditional scheme used to incorporate automation and automated functions into human operation relies on task division. This frequently relocates the human from an active controller to a supervisor, monitoring the system. In the case of human drivers interacting with vehicles equipped with automated driving technology, a more collaborative approach to human-system interaction has been taken [4]. At lower levels, this collaboration can be seen through the implementation of driver monitoring systems (DMS) and other safety warnings alerting or requesting the driver to take action to minimize risk. As the Level of Automation (LoA) increases, humans are expected to no longer intervene in driving tasks as drivers, but rather as passengers requesting emergency stops or remote operators supervising vehicle behavior [13]. The specific nature of the HAT's collaboration will depend on the LoA of the autonomous agent and the role envisioned for the human in each use case, such as applications in passenger or goods transport, or personal vehicles [14].

As regular human teams, HATs are target-driven, and task allocation plays an important role in the team's performance for both conditional (L3) and highly automated vehicle (L4) contexts. Further, both team members, human and ADS, can be interpreted as having distinct rules of behavior, tasks, and goals, both are affected by internal and external factors. Different shared control schemes, task allocation and control transition mechanisms are influenced by design and dynamic traffic environments [15]–[17]. Many factors have been studied in driving simulator experimental environments, exploring the effects of time pressure, stress, and sleep deprivation on driver takeover performance, or the effect automation has on the driver's behavior, attention, and cognitive load under long-term driving or emergency disengagements [18], [19], among other factors. A key challenge in the shared control scheme of L3 ADS is resolving conflicting assessments between the system and the driver during takeover events, particularly in the case of automation failure or the driver's lack of adequate situational awareness [16]. Similar challenges transfer over to L4 ADS driverless vehicles whose operation is supervised by remote operators, with the additional pressures of wireless communication reliability and impaired perception [20].

This work focuses on the case of a consumer-lever driver in a personal ADS-equipped vehicle, where the main functions and information flows are shown in Figure 1. In this configuration, both team members, the on-board driver and the ADS, can interact with the driving environment and other road users ('World') with their own perception and localization, DDT planning, and vehicle control functions [21]. The communication interface between them is expected to accommodate visual, audio, or haptic messages indicating the driver, vehicle, and automation status. A key interaction between team members is control transitions under nominal or emergency situations [22]. Control transitions or 'takeover' events have been studied from multiple perspectives, exploring the effect of external stimuli, human-system interface (HSI) designs, different time budgets, and scenario complexity, ultimately leading to takeover quality [2].

Model-based HRA approaches provide an advantageous opportunity to incorporate these human-system collaboration factors into the risk analysis of ADS-equipped vehicles. Leveraging decades of human performance model development in complex systems, this work seeks to translate the factors affecting the driver-ADS team's performance into PSFs. Developing a common terminology for both human and machine agents allows to conduct task analysis up to the same level of I-D-A phase detail [23]. Indeed, extending PSF nomenclature to describe driver-ADS team relationships paves the way for exploiting model-based PRA and HRA frameworks in shared-autonomy contexts. Multiple PSFs are employed in HRA models to express contextual, social, cognitive, and design factors' effect on human performance for both qualitative analysis and HEP quantification contexts. For instance, the Team-centered IDAC (Tc-IDAC) expanded the IDAC individual operator models to explicitly focus on team dynamics [10]. This model explores team-level tasks directed at each I-D-A stage and incorporates an Error Management module, i.e., how team members detect, indicate, and correct individual and team errors. For this, Tc-IDAC expanded upon team-related PSFs such as team cohesiveness, coordination, communication, composition, and leadership, and incorporated factors affecting communication between teammates. These team PSFs arise from the interaction dynamics between teammates, which, layered upon each team member's individual factors, affect the overall team's performance when addressing a common task. The focus on team behaviors from Tc-IDAC provides an advantageous starting point to include additional details for the taxonomy presented by [7] and the current state-of-the-art in Phoenix HRA Methodology [9]. The use of advanced HRA models that bring cognitive science and a model-based approach aims to reduce variability and increase reproducibility in HRA [9]. The remainder of this section discusses the relevance and applicability of individual, scenario, and team PSFs to driver-ADS HATs.

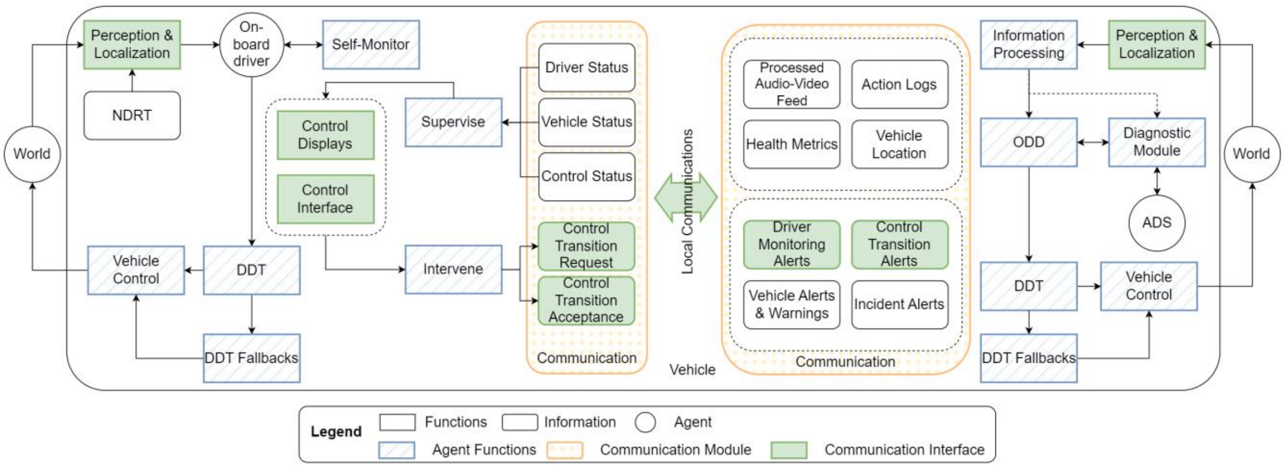

Figure 1. Diagram of Driver-ADS Team High-level Functions.

2.1. Individual Factors

Individual factors refer to internal factors affecting an agent's performance. These have historically been developed for human operators, hence, represent unobservable elements that require behavioral markers to be assessed, such as state of mind, temperament, intrinsic characteristics, physical and mental fitness, as well as the suitability of the individual to perform a certain task. Extending this concept to autonomous machine agents, i.e., an element in the system capable of independent decision-making, it is of interest to determine which may be applicable to express the underlying ADS software design.

- **Attention**: This factor refers to how cognitive resources are distributed by the agent to perform their tasks, usually divided into (a) attention to task and (b) attention to surroundings (e.g., alarms, team members). Driver attention has been thoroughly explored through subjective assessments and physiological metrics [24], [25]. In the ADS case, this concept can represent how system electrical/electronic (E/E) architecture design choices determine resource allocation to perform sensing, processing, planning, and execution tasks. Resource allocation is usually developed at early design stages, focusing on complying with safety and non-safety requirements [26], such as timing, redundancy, and freedom from interference (FFI) [27].
- **Physical & psychological abilities (PPA)**: This group of factors refer to the internal resources of the agent available to perform their tasks, including the level of alertness, fatigue, impairment, or relevant physical attributes. All these are of particular interest for impaired and distracted driving analysis [28], [29]. A similar concept may be extended to the ADS *Resource Availability* to perform tasks, such as system degradation (including elements of sensor or software reliability and calibration) and computational resource depletion (energy, memory, and processing capacity) during operation, significantly influenced by environmental factors, such as weather and lighting conditions [30]. A major factor to consider is the compatibility between existing hardware with over-the-air software and how this would affect operational assumptions [31].
- **Knowledge/Experience**: This umbrella term refers to the agent's understanding and knowledge accumulation about the system's design and operation, gained through training and interactions with the system, and affected by the individual's characteristics and variability [2], [32]. Currently, the driver relies almost entirely on their personal risk tolerance and driving experience (with and without ADAS/ADS technology) as receiving any specific training is unlikely [33]. For the ADS, this factor may be interpreted as *System Maturity*, encompassing multiple aspects, such as (1) data exposure (dynamic environmental conditions, HD maps accuracy, complexity of driving behavior models), (2) testing performance (real-world, closed tracks, simulation), and (3) the extent of its ODD expansion [34].
- **Skills**: This term represents the levels of perception (object detection, localization, tracking, hazard detection), decision-making (planning path, planning maneuvers, risk assessment), and control skills (vehicle control precision and responsiveness) of both agents [25], [35]. Additionally, a subfactor could be introduced to specifically address communication skills, i.e., both agents' ability to communicate effectively through HSI mechanisms. To assess this skill, the effect of elements such as decision transparency, explainability, and responsiveness on alerts, warnings, and control transitions events should be considered [36].
- **Bias**: This factor refers to the tendency of decision-making based on selected information, while excluding information that does not agree with the conceived conclusions. For the driver, multiple types of cognitive biases may be relevant (e.g., expectation, confirmation, belief). Given the lack of formal training, this factor becomes more important in determining the overall behavior of the driver [37]. For the ADS, this factor may express elements influenced by quality of the training data (data-based) and sophistication of the perception-localization-planning capabilities (algorithm-based bias), as well as sensor limitations under different conditions (sensor-based bias).
- **Perceived Familiarity**: This factor represents how the agents' decision-making process is influenced by recent events. In the case of the driver, the emphasis is placed on the memory of recent situations, while for the ADS this would be linked to the completeness and robustness of the training and testing phases. This factor is particularly important when assessing the driver's performance over multiple exposures to takeover events [19]. This factor may also be used to assess how operational experience is gradually introduced into the ADS's behavior through software and training data updates [15], [31].
- **Risk Tolerance**: This term encapsulates the agent's concern for safety and attitude towards risk. This is particularly important for rule-based decision making, traffic law compliance, and prioritizing safety based on contextual cues. For the driver, this may a be a defining factor in calibrating time budgets for takeovers and other emergency situations [38]. In the case of the ADS, the underlying risk tolerance is a result of the learnt driving behavior and imposed safety margins determining the level of 'aggressiveness' navigating uncertain circumstances [39].

2.2. Situation Factors

Situation or stressor-based factors are characteristics of the scenario external to the system and their effect on the agent's performance. In general, these factors are applicable to both the driver and the ADS. These factors are usually dynamic in nature, evolving as the scenario progresses, however some may only be set during design-time in the case of the ADS.

- **Conditioning events**: This factor refers to external events, latent failures, and alert-triggering events that influence scenario evolution, such as ODD breaches, vehicle failures, connectivity failures, and other factors that should trigger a DDT fallback response from the ADS and appropriate takeover actions from the driver [32]. Additionally, more subtle factors may be considered as well, such as less-than-adequate hardware or software conditions that affect system performance (e.g., faulty sensors, a flat tire). Similarly, conditioning events may also originate from the driver's behavior (as perceived by onboard DMS) [2].
- **External environment**: This factor can represent how both agents perceive the overall complexity of the driving conditions (e.g., weather, road geometry, traffic density) and interactions with other road users (e.g., other vehicles, cyclists, pedestrians). The effect of these elements may be mainly observed on the agent's risk tolerance, time-budget perception, and takeover performance [40].
- **Information load**: This represents the information and cues presented by elements external to the agent, including their team members and the external environment. Usually, this is used to represent the alarms and indicators. From the driver's perspective, the design and calibration of safety alarms plays a critical role in correctly assessing risk, especially in emergency situations [41]. To the ADS, this factor is related to system design factors (*Resource Availability*), where time and capacity limitations of conflict-resolving algorithms in data processing, localization, and planning tasks may lead to unreasonable latency and unsafe action execution. This may highly relevant when considering V2X applications, i.e., communications between vehicles and infrastructure [42].
- **Non-task load**: In this context, non-task related loads represent any tasks additional to the required DDTs. For the driver, this refers to the Non-Driving Related Tasks (NDRT) voluntary engagement and its effect on maintaining adequate situational awareness [28], [41]. For the ADS, this is directly affected by the system's design philosophy, e.g., the freedom from interference (FFI) between safety-critical and non-safety-critical tasks [26].
- **Task load**: This factor is an individual-level representation of the overall system design, accounting for the actual task allocation schemes and resulting task demand assigned to each agent [43], [44]. For the driver, this may be reflected in the procedure to takeover or handover control from/to the vehicle, including confirmation or veto actions, or the system's degree of reliance on their monitoring capacities. For the ADS, this factor reflects the demand placed upon its computational resources to manage competing interests, such as navigation and safety goals, potentially leading to delayed processing, performance degradation and potentially unsafe situations [27], [39].
- **Task complexity**: This factor refers to the cognitive and execution demands of the task at hand. This considers elements such as the difficulty in diagnosis, executing, knowledge required, procedural steps, precision required and the ambiguity of driving situations [43], [45]. As for *task load*, it can be extended to the computational demands of the ADS agent – for instance, differentiating between tasks performed within the ODD or in response to an ODD breach – and other contextual information increasing vehicle maneuver difficulty (e.g., road geometry, traffic density, weather conditions).
- **Time load**: This factor is defined as the ratio of time available to perform an action and the time to take an action. It is highly relevant to both the driver and the ADS agent, bounded by system design, physical limitations and time required to complete certain tasks [38], [46].
- **Stress**: This factor is considered solely for the driver. Previous studies have shown mixed results in terms of increasing or decreasing driving stress, indicating that the stress perceived by drivers in automated driving settings is greatly influenced by attitude towards driving and the level of driving automation [45].
- **Perceived situation severity/urgency**: This factor is related to the potential consequence perception of the agents and its influence on the decision-making process. Both severity and urgency play an important role for the driver [38], [47]. In the case of the ADS, however, these are expressed through the aforementioned risk metrics and safety margins that guide the decision-making process [17].
- **Perceived decision responsibility**: This factor is related to the agent's perception of the responsibility and accountability of their actions. This is particularly important to consider in the case of driving environments, where both agents interact with multiple other road users. This factor represents the

driver's perceived responsibility to intervene in a scenario should the ADS or themselves fail to prevent or mitigate a hazardous scenario [48].

2.3. Team Factors

Teams are complex and dynamic systems, whose performance is significantly affected by individual and situation factors described above. On one hand, teams increase problem-solving resource availability and provide the capacity to adapt to different situations. On the other, team dynamics introduce coordination challenges dependent on individual team-behaviors, such as cohesion and role awareness, communication, and authority gradients within the team [7], [49]. In turn, these elements are affected by the technical and non-technical skills of each member, as well as the mutual trust and understanding [4]. While additional granularity of factors can provide greater flexibility [10], the main team factors are summarized as the following:

- **Team Cohesion**: This factor refers to how team members interact with each other. This usually refers to social cohesion – understood as a measure of the effect of an individual's comfort within a team, degree of compatibility between team members, group morale and group attitude on team performance [37]. For the driver, it is important to characterize the comfort of the human operator interacting with the ADS as a team, reflected on their personal experience, the belief in shared goals (i.e., safety) and trust towards the autonomous agent [50], [51]. This social construct of cohesion may appear challenging to associate with ADS functions and performance. However, together with other risk behavior factors, these are determined by the ADS's design and training. In particular, given the rise of DMS and their relevance to system-initiated takeover requests, allowing a degree of calibration in detection and communication tasks may prove to be crucial while drivers gain more experience [29], [52].
- **Role Awareness**: Refers to how each team member perceives their duties, responsibilities, and their role within the team. It reflects multiple individual-level factors, and it is related to how each member acts in accordance with the expectations of the role. This factor may be crucial in automated driving contexts, as the attitude of the driver towards the team's roles, responsibilities, and goals may be critical under emergency situations [44]. A lack of understanding of the ADS's limitations and the implications of the driver's role could severely impact the team's ability to reach the high-level goals [48], [50].
- **Direct Supervision**: This factor represents the direct effect of leadership over a team's behavior, in which the individual factors of the leader are highly significant [10]. This factor plays a key role in team settings, particularly in those procedure-based and control room environments. In the context of automated driving, driver-ADS relationships are perceived as a vertical hierarchy, where the driver should have complete veto power over the ADS functions [53]. However, shared control paradigms and safety mechanisms against driver impairment do raise questions on whether the ADS can implement control transition actions without the driver's consent. Further discussion on this topic is required to uncover complex authority and error management dynamics within HATs [16], [54].
- **Team Communication**: This refers to the ability of team members to transfer and receive information to perform their tasks. Although frequently considered a system factor, automated driving contexts may warrant explicitly considering the design and availability of HSI mechanisms as communication-related team factors. Two distinct aspects can be considered: the driver's interaction with the HSI for non-safety related functions and the ADS's vehicle control transition management [36], [55]. From the ADS's perspective, the transmitted and received communication format, mode, and content are determined by design. Therefore, only aspects related to the quality and effectiveness of the communication are considered as team communication factors [4].
- **Team Coordination**: This factor serves as an output to characterize the team's performance of several tasks [10]. It involves the division of responsibilities and teamwork in planning, scheduling, and action implementation; hence, it involves all other team factors, and can be expressed through responsiveness and engagement indicators [49], [53]. This factor aims to represent the collaboration and task interdependence between the ADS and the driver, particularly in control transition scenarios [4], [17].

3. DISCUSSION

The safety and traffic effects of the projected increase of ADS technology's adoption on public roads continue to be actively researched. Assessing driver-ADS relationships as teams introduces a wide breadth of tools, literature, and technical language that can provide greater interpretability of the effect of the ADS design on the driver's performance, as opposed to characterizing them only as technology users [56]. Likewise, extending PSF nomenclature to describe driver-ADS team relationships paves the way for exploiting model-based PRA and HRA frameworks in mobility contexts. The use of quantitative model-based risk assessments can provide actionable information for the design of risk-reduction operational and design measures. For this, efforts are required to systematically collect data to quantify error and failure probabilities for hardware, software, and human elements in the system. In this sense, the PSF hierarchy introduced in [7] is intended to be used in a data-informed way, i.e., collapsible depending on data availability. While many of the identified factors related to the ADS are dependent on design-time decisions which may prove a challenge to quantify (e.g., risk tolerance, attention, system maturity, skills), there are many indicators that can be useful to track operation-time performance, efficiency, and latency (e.g., resource availability, task load). Characterizing driver and team factors can be more straightforward, given that the effect of system design on human-vehicle interactions have been extensively researched through human factors perspective. This presents many opportunities to develop and quantify metrics to incorporate in PRA models, such as responsiveness, rate of engagement, and overall takeover quality. For instance, scenario perception factors (e.g., perceived urgency, severity, responsibility) are mostly related to the subjective experience of the driver, assessed through well-established tools to measure workload [57], [58] and other methods based on physiological signals [40]. A highly studied factor is the driver's trust in the automated system; therefore, efforts should be directed to collecting and translating these studies' results into PRA data. Further discussion on potential data sources is required, particularly to represent the ADS design-time related factors. This work presents a discussion on the applicability of selected individual, scenario, and team factors to driver-ADS HATs. However, other elements, such as organizational and system factors, can play a critical role in this team's performance, considering the different organizations involved in the design and implementation of these systems. As consumer-level adoption increases, discussions about driver training and certification will also become more relevant [33], amidst ongoing regulatory discussions. Expanding the analysis to the systems that support the driver-ADS HAT also provides a path forward to fairly recognize the potential benefits of increasing ADAS/ADS deployment, for both driver performance and traffic safety, while also recognizing the effect of inadequate system design and calibration can lead to increased risk for road users.

4. CONCLUSION

The adoption of consumer-level vehicles equipped with automated driving functions will continue to be a highly researched topic. As in the case of other high-risk engineering systems, risk assessments have played a central role in regulation development. In the case of transportation environments, the short timeframes of traffic incidents highlight the importance of humans and autonomous machine systems working together as a team rather than isolated individual components. Establishing clear communication, trust calibration, and error correction mechanisms for on-board drivers-ADS teams is key to prevent or mitigate time-critical hazards., Focusing on team dynamics present in driver-ADS relationships through the technical language and models provided by HRA can help bridge a gap between extensive human factors studies, model-based risk assessments, and traffic safety assessments. As reliance on automated driving technology increases, understanding which factors influence takeover performance – and how these may be expressed in Risk Assessments – is an important step to assess risk and derive risk-reducing recommendations at design and operation time. The concepts presented in this work are the first foundations to develop an HRA model tailored for ADS-driver teams. This, with the purpose of providing input to qualitative and quantitative risk assessments, can then be used to develop risk reduction measures needed to support the safe operation of ADS, such as training, industry standards, and best practices.

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] T. A. O'Neill, C. Flathmann, N. J. McNeese, and E. Salas, "Human-autonomy Teaming: Need for a guiding team-based framework?," *Comput. Human Behav.*, p. 107762, Apr. 2023, doi: 10.1016/j.chb.2023.107762.
[2] A. D. McDonald *et al.*, "Toward Computational Simulations of Behavior During Automated Driving Takeovers: A Review of the Empirical and Modeling Literatures," *Hum. Factors J. Hum. Factors Ergon. Soc.*, vol. 61, no. 4, pp. 642–688, Jun. 2019, doi: 10.1177/0018720819829572.

[3] R. Gouribhatla and S. S. Pulugurtha, "Drivers' behavior when driving vehicles with or without advanced driver assistance systems: A driver simulator-based study," *Transp. Res. Interdiscip. Perspect.*, vol. 13, p. 100545, Mar. 2022, doi: 10.1016/j.trip.2022.100545.

[4] Y. Xing, C. Lv, D. Cao, and P. Hang, "Toward human-vehicle collaboration: Review and perspectives on human-centered collaborative automated driving," *Transp. Res. Part C Emerg. Technol.*, vol. 128, p. 103199, Jul. 2021, doi: 10.1016/j.trc.2021.103199.

[5] SAE International, "Taxonomy and Definitions for Terms Related to Driving Automation Systems for On-Road Motor Vehicles," *SAE Standard J3016_202104*, 2021.

[6] A. M. Boggs, R. Arvin, and A. J. Khattak, "Exploring the who, what, when, where, and why of automated vehicle disengagements," *Accid. Anal. Prev.*, vol. 136, p. 105406, Mar. 2020, doi: 10.1016/j.aap.2019.105406.

[7] K. M. Groth and A. Mosleh, "A data-informed PIF hierarchy for model-based human reliability analysis," *Reliab. Eng. Syst. Saf.*, vol. 108, pp. 154–174, Dec. 2012, doi: 10.1016/j.ress.2012.08.006.

[8] Y. H. J. Chang and A. Mosleh, "Cognitive modeling and dynamic probabilistic simulation of operating crew response to complex system accidents. Part 1: Overview of the IDAC Model," *Reliab. Eng. Syst. Saf.*, vol. 92, no. 8, pp. 997–1013, Aug. 2007, doi: 10.1016/j.ress.2006.05.014.

[9] N. J. Ekanem, A. Mosleh, and S. H. Shen, "Phoenix - A model-based Human Reliability Analysis methodology: Qualitative Analysis Procedure," *Reliab. Eng. Syst. Saf.*, vol. 145, pp. 301–315, Jan. 2016, doi: 10.1016/j.ress.2015.07.009.

[10] M. Azarkhil and A. Mosleh, "Dynamic behavior of operating crew in complex systems," *Proc. - Annu. Reliab. Maintainab. Symp.*, 2014, doi: 10.1109/RAMS.2014.6798446.

[11] M. C. G. S. P. Bandeira, A. R. Correia, and M. R. Martins, "General model analysis of aeronautical accidents involving human and organizational factors," *J. Air Transp. Manag.*, vol. 69, pp. 137–146, Jun. 2018, doi: 10.1016/j.jairtraman.2018.01.007.

[12] S. Farhadi, I. Mohammadfam, O. Kalatpour, and F. Ghasemi, "Determining performance shaping factors to assess human error in the emergency response team in chemical process industries: a case study," *Int. J. Occup. Saf. Ergon.*, vol. 29, no. 1, pp. 294–305, Jan. 2023, doi: 10.1080/10803548.2022.2043646.

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

[14] S. Wang, Z. Li, Y. Wang, W. Zhao, and H. Wei, "Quantification of safety improvements and human-machine tradeoffs in the transition to automated driving," *Accid. Anal. Prev.*, vol. 199, p. 107523, May 2024, doi: 10.1016/J.AAP.2024.107523.

[15] F. Warg, M. Skoglund, and M. Sassman, "Human Interaction Safety Analysis Method for Agreements with Connected Automated Vehicles," in *2021 IEEE 94th Vehicular Technology Conference (VTC2021-Fall)*, Sep. 2021, vol. 2021-Septe, pp. 01–07. doi: 10.1109/VTC2021-Fall52928.2021.9625202.

[16] F. Vanderhaegen, "Heuristic-based method for conflict discovery of shared control between humans and autonomous systems - A driving automation case study," *Rob. Auton. Syst.*, vol. 146, p. 103867, Dec. 2021, doi: 10.1016/j.robot.2021.103867.

[17] M. Marcano, S. DIaz, J. Perez, and E. Irigoyen, "A Review of Shared Control for Automated Vehicles: Theory and Applications," *IEEE Trans. Human-Machine Syst.*, vol. 50, no. 6, pp. 475–491, Dec. 2020, doi: 10.1109/THMS.2020.3017748.

[18] C. A. DeGuzman, D. Kanaan, S. A. Hopkins, and B. Donmez, "Takeover request (TOR) effects during different automated vehicle failures," *J. Intell. Transp. Syst.*, pp. 1–31, Mar. 2021, doi: 10.1080/15472450.2021.1891536.

[19] A. R. Samani, S. Mishra, and K. Dey, "Assessing the effect of long-automated driving operation, repeated take-over requests, and driver's characteristics on commercial motor vehicle drivers' driving behavior and reaction time in highly automated vehicles," *Transp. Res. Part F Traffic Psychol. Behav.*, vol. 84, pp. 239–261, Jan. 2022, doi: 10.1016/j.trf.2021.10.015.

[20] F. Tener and J. Lanir, "Driving from a Distance: Challenges and Guidelines for Autonomous Vehicle Teleoperation Interfaces," in *Conference on Human Factors in Computing Systems - Proceedings*, Apr. 2022, pp. 1–13. doi: 10.1145/3491102.3501827.

[21] C. Correa-Jullian, M. Ramos, A. Mosleh, and J. Ma, "Exploring Human-Autonomy Teams in Automated Driving System Operations," in *2024 IEEE 4th International Conference on Human-Machine Systems (ICHMS), Toronto, Canada, 2024*.

[22] C. Correa-Jullian, M. Ramos, A. Mosleh, and J. Ma, "Task Allocation and Control Transitions in Autonomous Driving System Operations," in *34th European Safety and Reliability Conference (ESREL 2024), Cracow, Poland, 2024*, 2024, vol. 1.

[23] 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, p. 106697, Mar. 2020, doi: 10.1016/j.ress.2019.106697.

[24] J. Araluce, L. M. Bergasa, M. Ocaña, E. López-Guillén, R. Gutiérrez-Moreno, and J. F. Arango, "Driver Take-Over Behaviour Study Based on Gaze Focalization and Vehicle Data in CARLA Simulator," *Sensors*, vol. 22, no. 24, p. 9993, Dec. 2022, doi: 10.3390/s22249993.

[25] N. Zangi, R. Srour-Zreik, D. Ridel, H. Chasidim, and A. Borowsky, "Driver distraction and its effects on partially automated driving performance: A driving simulator study among young-experienced drivers," *Accid. Anal. Prev.*, vol. 166, p. 106565, Mar. 2022, doi: 10.1016/j.aap.2022.106565.

[26] International Organization for Standardization, "ISO 26262:2018, Road vehicles — Functional safety," 2018. https://www.iso.org/standard/68383.html

[27] H. Askaripoor, T. Mueller, and A. Knoll, "E/E Designer: a Framework to Design and Synthesize Vehicle E/E Architecture," *IEEE Trans. Intell. Veh.*, 2023, doi: 10.1109/TIV.2023.3324617.

[28] A. McKerral, K. Pammer, and C. Gauld, "Supervising the self-driving car: Situation awareness and fatigue during highly automated driving," *Accid. Anal. Prev.*, vol. 187, p. 107068, Jul. 2023, doi: 10.1016/J.AAP.2023.107068.

[29] S. Ayas, B. Donmez, X. Tang, and C. Author, "Drowsiness Mitigation Through Driver State Monitoring Systems: A Scoping Review," *Hum. Factors J. Hum. Factors Ergon. Soc.*, vol. 0, no. 0, pp. 1–26, Nov. 2023, doi: 10.1177/00187208231208523.

[30] Y. Zhang, A. Carballo, H. Yang, and K. Takeda, "Perception and sensing for autonomous vehicles under adverse weather conditions: A survey," *ISPRS J. Photogramm. Remote Sens.*, vol. 196, pp. 146–177, Feb. 2023, doi: 10.1016/J.ISPRSJPRS.2022.12.021.

[31] X. Zhao, K. Salako, L. Strigini, V. Robu, and D. Flynn, "Assessing safety-critical systems from operational testing: A study on autonomous vehicles," *Inf. Softw. Technol.*, vol. 128, p. 106393, Dec. 2020, doi: 10.1016/j.infsof.2020.106393.

[32] S. Mishler and J. Chen, "Effect of automation failure type on trust development in driving automation systems," *Appl. Ergon.*, vol. 106, p. 103913, Jan. 2023, doi: 10.1016/J.APERGO.2022.103913.

[33] S. M. Casner and E. L. Hutchins, "What Do We Tell the Drivers? Toward Minimum Driver Training Standards for Partially Automated Cars," *J. Cogn. Eng. Decis. Mak.*, vol. 13, no. 2, pp. 55–66, Jun. 2019, doi: 10.1177/1555343419830901/ASSET/IMAGES/LARGE/10.1177_1555343419830901-FIG1.JPEG.

[34] S. Yang, A. Mckerral, M. D. Mulhall, M. G. Lenné, B. Reimer, and P. Gershon, "Takeover Context Matters: Characterising Context of Takeovers in Naturalistic Driving using Super Cruise and Autopilot," in *ACM International Conference Proceeding Series*, Sep. 2023, vol. 11, pp. 112–122. doi: 10.1145/3580585.3606459.

[35] X. Di and R. Shi, "A survey on autonomous vehicle control in the era of mixed-autonomy: From physics-based to AI-guided driving policy learning," *Transp. Res. Part C Emerg. Technol.*, vol. 125, p. 103008, Apr. 2021, doi: 10.1016/j.trc.2021.103008.

[36] A. M. Tinga *et al.*, "Development and evaluation of a human machine interface to support mode awareness in different automated driving modes," *Transp. Res. Part F Traffic Psychol. Behav.*, vol. 92, no. April 2022, pp. 238–254, Jan. 2023, doi: 10.1016/j.trf.2022.10.023.

[37] J. B. Manchon, M. Bueno, and J. Navarro, "How the initial level of trust in automated driving impacts drivers' behaviour and early trust construction," *Transp. Res. Part F Traffic Psychol. Behav.*, vol. 86, pp. 281–295, Apr. 2022, doi: 10.1016/j.trf.2022.02.006.

[38] F. Tanshi and D. Soffker, "Determination of Takeover Time Budget Based on Analysis of Driver Behavior," *IEEE Open J. Intell. Transp. Syst.*, vol. 3, pp. 813–824, Nov. 2022, doi: 10.1109/OJITS.2022.3224677.

[39] S. S. Lodhi, N. Kumar, and P. K. Pandey, "Autonomous vehicular overtaking maneuver: A survey and taxonomy," *Veh. Commun.*, vol. 42, p. 100623, Aug. 2023, doi: 10.1016/J.VEHCOM.2023.100623.

[40] J. Ma, Y. Wu, J. Rong, and X. Zhao, "A systematic review on the influence factors, measurement, and effect of driver workload," *Accid. Anal. Prev.*, vol. 192, p. 107289, Nov. 2023, doi: 10.1016/j.aap.2023.107289.

[41] T. Louw, J. Kuo, R. Romano, V. Radhakrishnan, M. G. Lenné, and N. Merat, "Engaging in NDRTs

affects drivers' responses and glance patterns after silent automation failures," *Transp. Res. Part F Traffic Psychol. Behav.*, vol. 62, pp. 870–882, Apr. 2019, doi: 10.1016/j.trf.2019.03.020.

[42] R. Xu, H. Xiang, Z. Tu, X. Xia, M. H. Yang, and J. Ma, "V2X-ViT: Vehicle-to-Everything Cooperative Perception with Vision Transformer," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, Mar. 2022, vol. 13699 LNCS, pp. 107–124. doi: 10.1007/978-3-031-19842-7_7.

[43] J. Stapel, F. A. Mullakkal-Babu, and R. Happee, "Automated driving reduces perceived workload, but monitoring causes higher cognitive load than manual driving," *Transp. Res. Part F Traffic Psychol. Behav.*, vol. 60, pp. 590–605, 2019, doi: 10.1016/j.trf.2018.11.006.

[44] S. Khastgir, S. Birrell, G. Dhadyalla, and P. Jennings, "Effect of Knowledge of Automation Capability on Trust and Workload in an Automated Vehicle: A Driving Simulator Study," in *Advances in Intelligent Systems and Computing*, 2019, vol. 786, pp. 410–420. doi: 10.1007/978-3-319-93885-1_37.

[45] D. D. Heikoop, J. C. F. de Winter, B. van Arem, and N. A. Stanton, "Acclimatizing to automation: Driver workload and stress during partially automated car following in real traffic," *Transp. Res. Part F Traffic Psychol. Behav.*, vol. 65, pp. 503–517, Aug. 2019, doi: 10.1016/j.trf.2019.07.024.

[46] C. Marberger, H. Mielenz, F. Naujoks, J. Radlmayr, K. Bengler, and B. Wandtner, "Understanding and applying the concept of 'Driver availability' in automated driving," in *Advances in Intelligent Systems and Computing*, 2018, vol. 597, pp. 595–605. doi: 10.1007/978-3-319-60441-1_58.

[47] S. Becker, S. Brandenburg, and M. Thüring, "Driver-initiated take-overs during critical braking maneuvers in automated driving – The role of time headway, traction usage, and trust in automation," *Accid. Anal. Prev.*, vol. 174, p. 106725, Sep. 2022, doi: 10.1016/j.aap.2022.106725.

[48] H. Azevedo-Sa, H. Zhao, C. Esterwood, X. J. Yang, D. M. Tilbury, and L. P. Robert, "How internal and external risks affect the relationships between trust and driver behavior in automated driving systems," *Transp. Res. Part C Emerg. Technol.*, vol. 123, p. 102973, Feb. 2021, doi: 10.1016/j.trc.2021.102973.

[49] D. H. Ham, W. J. Jung, and J. Park, "Identifying key factors affecting the performance of team decision-making based on the analysis of investigation reports issued from diverse industries," *Reliab. Eng. Syst. Saf.*, vol. 206, no. November 2020, p. 107304, 2021, doi: 10.1016/j.ress.2020.107304.

[50] H. Pan, K. Xu, Y. Qin, and Y. Wang, "How does drivers' trust in vehicle automation affect non-driving-related task engagement, vigilance, and initiative takeover performance after experiencing system failure?," *Transp. Res. Part F Traffic Psychol. Behav.*, vol. 98, pp. 73–90, Oct. 2023, doi: 10.1016/J.TRF.2023.09.001.

[51] W. Payre *et al.*, "Understanding Drivers' Trust After Software Malfunctions and Cyber Intrusions of Digital Displays in Automated Cars," *Hum. Factors Transp.*, vol. 60, pp. 320–328, 2022, doi: 10.54941/AHFE1002463.

[52] Y. Wang, W. Zhang, and R. Zhou, "Speech-based takeover requests in conditionally automated driving: Effects of different voices on the driver takeover performance," *Appl. Ergon.*, vol. 101, p. 103695, May 2022, doi: 10.1016/j.apergo.2022.103695.

[53] J. B. Lyons, K. Sycara, M. Lewis, and A. Capiola, "Human–Autonomy Teaming: Definitions, Debates, and Directions," *Front. Psychol.*, vol. 12, p. 1932, May 2021, doi: 10.3389/fpsyg.2021.589585.

[54] A. Bhardwaj *et al.*, "Who's the boss? Arbitrating control authority between a human driver and automation system," *Transp. Res. Part F Traffic Psychol. Behav.*, vol. 68, pp. 144–160, Jan. 2020, doi: 10.1016/j.trf.2019.12.005.

[55] H. Muslim, C. Kiu Leung, and M. Itoh, "Design and evaluation of cooperative human–machine interface for changing lanes in conditional driving automation," *Accid. Anal. Prev.*, vol. 174, p. 106719, Sep. 2022, doi: 10.1016/j.aap.2022.106719.

[56] J. C. Walliser, E. J. de Visser, E. Wiese, and T. H. Shaw, "Team Structure and Team Building Improve Human–Machine Teaming With Autonomous Agents," *J. Cogn. Eng. Decis. Mak.*, vol. 13, no. 4, pp. 258–278, Dec. 2019, doi: 10.1177/1555343419867563.

[57] A. Pauzié, "A method to assess the driver mental workload: The driving activity load index (DALI)," *IET Intell. Transp. Syst.*, vol. 2, no. 4, p. 315, 2008, doi: 10.1049/iet-its:20080023.

[58] S. G. Hart and L. E. Staveland, "Development of NASA-TLX (Task Load Index): Results of Empirical and Theoretical Research," *Adv. Psychol.*, vol. 52, no. C, pp. 139–183, Jan. 1988, doi: 10.1016/S0166-4115(08)62386-9.