From Awareness to Intent: Mitigating Silent Driving System Failures through Prospective Situation Awareness Enhancing Interfaces

Jiyao Wang*
jwanggo@connect.ust.hk
The Hong Kong University of Science
and Technology (Guangzhou)
Guangzhou, Guangdong, China

Song Yan*
syan931@connect.hkust-gz.edu.cn
The Hong Kong University of Science
and Technology (Guangzhou)
Guangzhou, Guangdong, China

Xiao Yang
xyang856@connect.hkust-gz.edu.cn
The Hong Kong University of Science
and Technology (Guangzhou)
Guangzhou, Guangdong, China

Qihang He
e1538701@u.nus.edu
National University of Singapore
Singapore

Ange Wang
awang324@connect.hkust-gz.edu.cn
The Hong Kong University of Science
and Technology (Guangzhou)
Guangzhou, Guangdong, China

Chenglin Liu
cliu549@connect.hkust-gz.edu.cn
The Hong Kong University of Science
and Technology (Guangzhou)
Guangzhou, Guangdong, China

Chenglin Chen
cchen363@connect.hkust-gz.edu.cn
The Hong Kong University of Science
and Technology (Guangzhou)
Guangzhou, Guangdong, China

Zhenyu Wang
zwang209@connect.hkust-gz.edu.cn
The Hong Kong University of Science
and Technology (Guangzhou)
Guangzhou, Guangdong, China

Dengbo He[†]
dengbohe@hkust-gz.edu.cn
The Hong Kong University of Science
and Technology (Guangzhou)
Guangzhou, Guangdong, China

Normal Operation   Silent Failure

Figure 1: Illustration of the difference between an automated driving system (ADS) with normal operation and the silent failures. In case of failures, the ADS continues to control the vehicle, but its internal representation of the situation is misaligned with the actual road environment: either the hazard is not detected at the perception layer and/or the planned trajectory conflicts with the hazard at the planning/decision layer, and no TOR is issued in case of all failures.

Abstract

Silent automation failures, where a system fails to detect a hazard without warning, pose a critical safety challenge for partially automated vehicles. While research has mostly focused on takeover requests, how to support a driver in silent failure remains underexplored. We conducted a multi-modal driving simulator study with 48 participants to investigate how different Prospective Situation Awareness Enhancement (PSAE) interfaces, delivered via augmented reality head-up display, affect takeover performance. By integrating behavioral, subjective psychological, and physiological data, our analysis suggests that situational awareness (SA) serves as an important moderating factor through which PSAE interfaces improve takeover performance. Further, we found that providing perceptual cues was most effective in enhancing SA, while communicating system intent was superior for building trust. Finally, we identified a potential correlate of SA in the neuroactivity. Overall, contributes to understanding how transparency-oriented interfaces may support drivers and provides design insights into HMI design for silent failures.

*Both authors contribute equally.
[†]Corresponding author

CCS Concepts

• **Human-centered computing → Empirical studies in HCI**.

Keywords

semi-automated vehicles; silent failure; prospective situation awareness enhancement; situational awareness; empirical study; quantitative analysis

1 Introduction

The increasing integration of automation into daily life fundamentally changes the nature of human-technology interaction [41]. A core challenge in this evolving relationship is designing for transparency and trust in a shared-control paradigm, especially when a human operator must remain "in the loop." In the context of partially automated vehicles (SAE Level 2) [19], which are gradually dominating the consumer market [46], this is not just a user experience problem but a critical safety issue. At this level, human drivers are still responsible for monitoring the environment and responding to system limitations, either actively or prompted by takeover requests (TORs) from automated driving systems (ADS).

Existing human-ADS interaction research predominantly focused on TOR-initiated takeover scenarios [43, 82, 83, 92]. However, silent failures without TORs are likely more common and safety-critical in SAE Level 2 vehicles. Hazard denotes any on-path road user or object whose projected motion would conflict with the ego vehicle in the absence of driver intervention [104]. As shown in Figure 1, the inability of ADS to detect and respond to road hazards poses a severe challenge that has not yet been fully addressed in existing research. In these highly safety-critical situations, a system may encounter a failure without warning drivers in advance or until very late (e.g., due to an unforeseen malfunction), leaving the driver to detect and respond to the failure independently with a very short time buffer [48]. Additionally, the time budget and cognitive workload needed from drivers when they encounter hazards are shaped by how perceptually accessible those hazards are [84]. Previous studies [20, 104] distinguish two scenario types based on hazard visibility: visible hazards, which are fully observable as they materialize, and invisible hazards, which are initially hidden and must be detected via contextual cues. Invisible hazards typically demand earlier predictions and broader visual searches; while in ADS with silent failures, these differences in hazard visibility translate directly into distinct situation awareness requirements and takeover time budgets.

Some works [37, 50, 53] have investigated the use of explanations to enhance driver awareness, but these works assume the system's internal perception is aligned with the actual environment. In fact, this assumption breaks down in silent failure scenarios, where a malfunction in system perception can be the cause of system failure [22]. Despite a few studies [34, 47] have explored improving takeover performance in silent failure with different explanatory information, to our knowledge, little research has investigated how information regarding key ADS components, i.e., perception and maneuver planning, influences drivers' takeover decisions in silent failure events under different environmental conditions (e.g., lighting and hazard types), which would manipulate the complexity and difficulty of the event [45].

To address this gap, we investigate the efficacy of the prospective situation awareness enhancements (PSAE) [39] in silent failure scenarios. This form of information, fundamentally different from retrospective explanations that justify past system actions after a failure has occurred, is designed to enhance the driver's Level 1 (*Perception*) and Level 3 (*Projection*) SA [33] by offering a continuous, forward-looking stream of information about the system's perceptual state (what the system 'sees') and planned maneuver. Drawing on theories of embodied cognition and cognitive neuroscience [10], we posit that a driver's takeover behavior is not merely a reactive response but the culmination of a dynamic interplay between their psychological state, neuro-physiological activities, and the environmental context [40]. A driver's ability to respond to a hazardous event is contingent on the timely and coordinated interplay of mental processing and motor preparation [51]. In line with classical work on driving as hierarchical information processing [66] and with situation awareness theory [32], we assume that drivers must first perceive, comprehend, and project the state of the environment before they can select and execute appropriate actions [36, 62, 88]. In addition, indicators from neuro-physiological signals can provide continuous, objective correlates of attentional engagement and cognitive activities [64].

Therefore, based on dynamic augmented–reality head–up display (AR-HUD), we adopt a multi-layered framework, in which PSAE as an experimental variable, to empirically examine how its variation, in conjunction with ambient environmental factors, including lighting conditions (day vs. night) and hazard visibility (visible vs. invisible), influences drivers' SA, perceived safety, trust, EEG/ERSP and EMG-based measures, and ultimately their takeover performance [24]. This provides a more nuanced understanding of the perception and trajectory planning information of ADS may affect human-ADS cooperation in safety-critical scenarios. Specifically, we use electroencephalography (EEG) to analyze Event-Related Spectral Perturbations (ERSP) [59], a measure of oscillatory brain activity associated with cognitive processes, and electromyography (EMG) to track motor preparation [15]. From these, we measure the temporal alignment between the peak of a driver's cognitive processing and the onset of their motor preparation, thereby providing a unique insight into the temporal coupling of cognition and action [77] when drivers are provided different PSAE information in different silent failure scenarios.

To investigate these complex relationships and address the aforementioned research gaps, this study aims to answer the following research questions:

- **RQ1**: How do different PSAE designs and environmental factors (i.e., lighting and hazard visibility) influence drivers' self-reported Situation Awareness (SA), perceived safety, and

trust, as well as objective EEG- and EMG-derived physiological indicators during silent automation failures?

- **RQ2**: What is the intricate relationship between a driver's subjective psychological states (e.g., SA, perceived safety, trust) and their physiological indicators?
- **RQ3**: What are the underlying mechanisms through which PSAE interfaces impact takeover performance? Specifically, do psychological states and/or physiological indicators serve as the primary causal pathways?

Our contributions are as follows: (1) We quantitatively model the multifaceted impacts of PSAE interfaces, lighting, and hazard visibility on drivers' internal states (i.e., psychological states and physiological indicators) in the context of silent ADS failures. (2) We examine hierarchical pathways connecting driver psychology and physiology to takeover behavior, and our models suggest that, in this study, the effect of the interface can be indirectly mediated by drivers' SA. (3) We offer an initial mechanistic interpretation of how our intervention may influence safety-related outcomes in our simulator setting, and derive empirically grounded design implications for creating transparent and trustworthy automated systems in silent failures.

2 Related Work

2.1 Silent Failures in Automated Driving

SAE Level 2 places human drivers in a critical supervisory role, responsible for monitoring the environment and intervening when automation degrades or fails [19]. The majority of design-focused studies still center on TOR content, modality, and timing [31, 45, 54]. While the *silent failures* pose an equally, if not more, critical safety risk [48]. As highlighted in recent reviews and taxonomies of driving automation failures, such failures can arise at different stages of the automated driving pipeline, including perception, decision-making, and control [48]. Real-world crashes (e.g., Tesla/Florida, 2016; Uber/Tempe, 2018) illustrate how perception-level failures can escalate when drivers remain unaware of imminent hazards [3, 73]. Analytical work has found that such failures may arise at the perception layer of ADS [7] and that the failures in perception may occur even when downstream planning appears reasonable. Thus, in this work, we focus on a practically important subset of silent failures in SAE Level 2 systems, in which the ADS's perceived and/or planned state of the world is misaligned with the actual hazard on the road and therefore no TOR is issued. Operationally, this manifests as hazards that are not correctly marked in the perception layer and/or as planned trajectories that conflict with the hazard, which are made visible to the driver through the PSAE interfaces.

Beyond incident analyses, recent human-factors experiments have begun to reveal the risks of silent failures. Louw et al. [56] showed that, following silent lane-keeping failures, drivers engaged in non-driving-related tasks had slower responses and exhibited more lane excursions. Bianchi Piccinini et al. [2] reported delayed braking and degraded takeover quality in silent failure conditions relative to manual driving. Mole et al. [69] modeled takeover timing in silent failures and demonstrated how individual and situational factors systematically affect response latency. However, very few researchers have explored mitigation strategies to improve takeover performance in silent failures. For example, projecting planned trajectories on the windshield has been shown to improve takeover responses in silent failures [47], and AR-HUDs can assist drivers during malfunctions [34]. However, their limited types of PSAE and relatively simple hazardous scenarios are unable to comprehensively assess the support provided by different PSAE systems for driving safety in dynamic traffic environments when silent failures occur.

2.2 Interpretability and Explanations for Automated Driving Systems

Interpretability refers to an ADS's ability to provide meaningful and understandable information about its status, maneuvers, and uncertainties [28, 74]. As ADS increasingly rely on complex AI, their internal reasoning often diverges from human expectations, challenging user understanding and adversely affecting users' trust [67]. A wide range of explainable-AI approaches have been proposed to increase transparency in driving contexts, including attention/saliency visualizations [9, 21], interpretable perception outputs such as object detection and semantic segmentation [1, 85], planned trajectory visualizations [17, 75], and uncertainty representations that signal system confidence [65, 86].

Situation awareness (SA) is widely defined as "the perception of the elements in the environment within a volume of time and space, the comprehension of their meaning, and the projection of their status in the near future" [32]. In partially automated driving, HMI perception-oriented visualizations mainly support Level 1 SA ("what is there"), while explanations at the decision level relate more to how these perceptions are used for maneuvers (Level 3 SA, "what will happen next"). This view is consistent with classical hierarchical models of driving [66], where maintaining sufficient tactical-level SA is a prerequisite for safe operational control during a takeover. Most deployed or evaluated explanations at *decision-level* (e.g., 'what the car will do'), implicitly assumed that the perception is accurate. This assumption breaks down in *silent failure* scenarios, where the core problem is a misperception of the environment rather than a poor decision [22]. Recent works have started to focus on *perception-level* transparency: e.g., visualizing recognized elements in a scene or system predictions to improve drivers' SA in TOR-initiated scenarios [16, 17, 37]. Yet, how prospective, perception-level information (what the system 'currently sees' and 'expects to happen next') affects drivers' psychological readiness, physiological preparation, and ultimate takeover performance under silent failures remains underexplored.

Beyond the content in the displays, researchers have also explored how explanations can be delivered to drivers. HUD-based explanations, such as AR overlays of hazards or trajectories, can improve takeover quality [34]. Multimodal approaches, combining visual with auditory explanations, can further reduce cognitive load [31, 54]. However, while prior works have explored head–up display for providing planned trajectories or perceptual cues, they have largely focused on TOR scenarios or have not systematically investigated the combined effects of both perception and planned maneuver information in silent failure conditions.
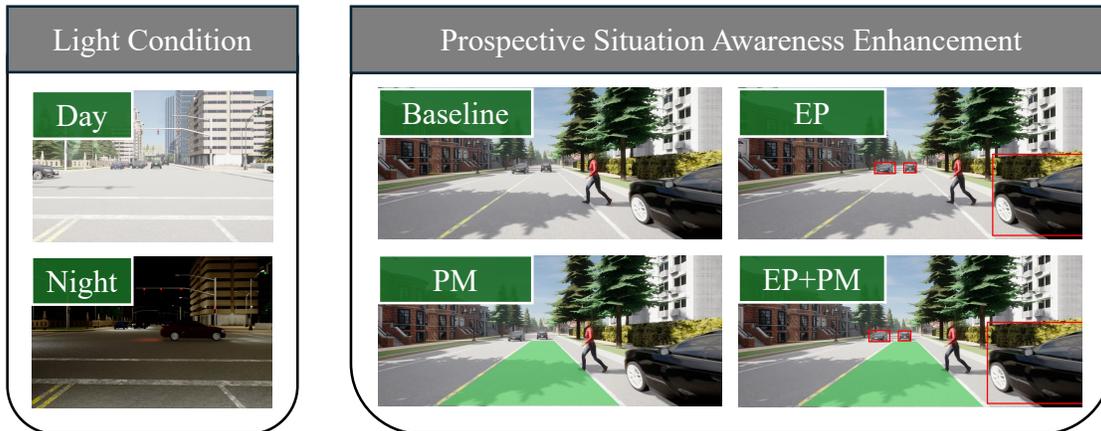
Figure 2: Illustration of the two between-subject conditions, including four PSAEs and two light conditions presented to the driver in the simulator.

2.3 This Study

To address the above-mentioned gaps, our study aims to examine how perception-level, forward-looking information, both alone and combined with environmental factors, can affect drivers' internal states and safety-critical behavior in L2 driving. Specifically, our work simulated a dynamic AR-HUD, offering continuous perception- and projection-level information to support drivers in silent failure scenarios, as user studies indicate that such explanations can bolster trust, SA, and takeover quality—especially when prompted by TORs [31, 45, 50, 54]. Further, given that lighting conditions, hazard visibility, and task workload have been shown to affect trust, attention allocation, and takeover performance [24, 92, 99], we further evaluated how drivers' psychological states (i.e., self-reported SA, perceived safety, and trust), their EEG- and EMG-based physiological measures, and the manipulated environmental factors (day/night and visible/invisible hazards) are dynamically coupled when different PSAE information is provided on an AR-HUD.

3 Methodology

3.1 Participants

We recruited 48 participants (24 male, 24 female) from the local community via online postings and campus mailing lists. Eligible participants must have held a valid driving licence for at least two years, own a vehicle with ADS or self-report to be familiar with ADS features (such as Autopilot in Tesla or equivalent systems), and have driven at least 5,000 km in the past year. Drivers' age ranged from 23 to 42 ($M = 27.54, SD = 4.60$), with years of licensure from 2 to 19 ($M = 6.52, SD = 3.57$). This population represents active, non-novice urban drivers who are more likely to be early adopters of SAE Level-2 automation [44, 72, 97]. Participants self-reported having normal or corrected-to-normal vision and no neurological or motor impairments. Each participant received 100 CNY for participating in this study. The study protocol was approved by the Human and Artefacts Research Ethics Committee at the Hong Kong University of Science and Technology (Guangzhou) (protocol: HSP-2023-0009).

3.2 Prospective Situation Awareness Enhancement

Our Prospective Situation Awareness Enhancement (PSAE) was designed with three guiding objectives: (1) *Transparency of perception*, i.e., what ADS detects, so drivers can notice perceptual misses; (2) *Action affordance*, which links perception to planned maneuvers so that drivers can quickly infer whether the system intends to avoid a detected object or not; (3) *Low intrusion & stability*, which provides continuous information that improves anticipation without producing attentional overload or distracting flicker [16, 17, 34].

Subsequently, except for the baseline condition (no PSAE applied), we implemented three PSAE visualization strategies, each rendered as dynamic overlays on the forward view (see Figure 2):

- **Environment Perception (EP).** ADS-detected objects are highlighted using color-coded bounding boxes and textual labels. Color mapping: *yellow* for traffic infrastructure (e.g., lights/signs), *blue* for pedestrians, and *red* for vehicles/obstacles. The numeric confidence score is not presented to avoid excessive cognitive load. [9, 49, 57, 79].
- **Planned Maneuver (PM).** The ADS's short-term planned trajectory is rendered as a semi-transparent green polyline extending from the ego-vehicle for a 3–5 s horizon, together with a shaded stopping envelope, indicating the current braking/steering safety margin. The trajectory rendering communicates intended lateral and longitudinal actions and visually shows potential future conflicts with overlaid objects [47, 75].
- **Combined (EP+PM).** Both EP and PM are presented concurrently, enabling drivers to directly map detected elements to the system's intended maneuver and thus quickly identify perception–action mismatches (e.g., a missing bounding box while the PM path intersects an undetected pedestrian).

Because the experiment studies *silent failures* (perceptual misses without a TOR), we implemented two types of overlays per traffic scenario. Normal operation: overlays faithfully reflect the ADS detections and planned trajectory recorded during the Wizard-of-Oz run (bounding boxes present, PM avoids hazards). Silent failure: overlays simulate perception misses by omitting expected

bounding boxes (EP miss) and/or showing a planned trajectory that intersects the hazard (PM mismatch). Importantly, in silent-failure trials, *no* takeover request or alarm is presented; PSAE is the only feedforward information available to the driver. All visual elements were designed to maintain legibility and minimize clutter under both day and night conditions.

**Table 1: Categories of silent-failure scenarios in the study**

Hazard visibility	Scenario description	Illustration
Invisible	Scenario 1: Left-turning at an unsignalized junction with the view obstructed by trees and fences and roadside objects (visible ∼50 m), indicating possible cross-traffic.	
	Scenario 5: Pedestrian stepping into the road from behind an advertising board on a curved road; the board (visible ∼60 m) signals possible hidden pedestrians.	
	Scenario 6: Pedestrian emerging from behind a parked vehicle when ego car goes straight; the parked car (visible ∼70 m) serves as the cue.	
Visible	Scenario 2: Pedestrian suddenly crossing from the right at a signalized intersection against the red light.	
	Scenario 3: Parked car in the horizon on the right pulling out and merging abruptly into the ego lane.	
	Scenario 4: Encountering a road construction zone while driving straight ahead.	

Note: The yellow car indicates the ego vehicle with green arrows showing its intended path. Hazards are represented in red: continuous arrows for ongoing motion and triangles for the onset of movement.

3.3 Experimental Design and Procedures

The study employed a mixed 4×2×2 factorial design with one within-subject factor and two between-subject factors. The between-subject factors were PSAE type (Baseline, EP, PM, and EP+PM) and lighting condition (day vs. night). The within-subject factor was hazard visibility (visible vs. invisible hazards). The six silent-failure events consisted of three scenarios with visible hazards and three scenarios with invisible hazards (Table 1). Each participant experienced all six events exactly once, under a single PSAE type and lighting condition, such that every driver encountered three visible-hazard and three invisible-hazard failures in one continuous loop drive.

Hazard visibility. Following [104], visible hazards are directly observable in the driver's forward field of view when they enter the scene (e.g., a pedestrian already in view crossing against a red light, a clearly marked construction zone, or a parked vehicle that is continuously visible before pulling out). In contrast, invisible hazards are initially occluded by other objects and must be inferred from contextual cues (e.g., a pedestrian emerging from behind an advertising board or a parked vehicle, or cross-traffic hidden behind roadside structures). To give participants enough time to develop situation awareness, all contextual cues (for invisible hazards) or hazard visibility horizons (for visible hazards) are from 8 to 15

(a) Driving simulator of the experiment

(b) Bird's-eye view of the route
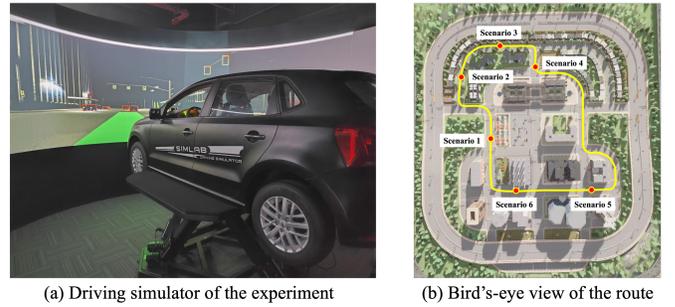
Figure 3: The demonstration of (a) six-degree-of-freedom driving simulator and (b) experimental route setup.

seconds, exceeding the recommended minimum of 5 seconds [68]. The six events were placed at fixed locations along a closed urban route in CARLA Town 5 (Fig. 3(b)) [29]. Scenario order was counter-balanced within each PSAE × lighting group using a Latin-square scheme by varying the starting point on the loop, such that each scenario appeared equally often in each ordinal position across the six drivers in a group.

Lighting. We implemented two ambient lighting conditions in CARLA, corresponding to a clear daytime and a nighttime urban environment. In the day condition, the simulation time-of-day was set to midday with high ambient illumination and no artificial street lighting. In the night condition, the time-of-day was set to evening with low ambient illumination, active street lights, and ego-vehicle headlights, while keeping road geometry, traffic, and hazard timings identical to the day scenarios. These settings ensured that only the lighting conditions, rather than the underlying traffic dynamics, differed between day and night.

Before driving, participants were invited based on the screening questionnaire (including demographics and driving history information) and first completed informed consent. To mitigate the potential influence of different ADS usage experience on drivers' responses to silent failures, they then received standardized instruction and completed a practice/demo drive containing one practice silent-failure event (not used in the formal experiment) to familiarize them with the manual driving of the vehicle, how to take over the control of the vehicle, and the assigned PSAEs. During the experiment, takeover was detected when a participant initiated steering beyond ±10° or depressed the brake pedal beyond 10% travel, consistent with prior studies [55].

Then, the formal experiment started, which contained six silent-failure events. Whenever a takeover was detected throughout the drive, the simulator paused. Takeovers occurring outside the designated event windows (around 10 seconds from first visibility of the hazard/cue object to collision) were classified as error takeovers. If it happened, the drive resumed from the paused point. Successful takeovers caused the remainder of that failure segment to be skipped to simulate successful avoidance. We further extracted the *Takeover Lead Time* and *Takeover Success* for further analysis.

After each takeover or scenario end (if no takeover was detected in a scenario, which lasted around 5 minutes), participants completed a 10-item Situation Awareness questionnaire for assessing situation awareness [87], a TiAD trust scale [61], and a Perceived Safety scale adopted from [4].

3.4 Apparatus

Experiments were run on a six-degree-of-freedom motion-capable driving simulator (WIVW GmbH) with a realistic vehicle cabin (force-feedback steering, pedals). Visuals were presented via projection over a 210° horizontal field of view at a combined resolution of 5760×1200 pixels to create an immersive urban environment (Fig. 3). To ensure repeatability and precise control over silent failures, we adopted a Wizard-of-Oz approach: experimenters manually drove and recorded scenario runs in CARLA to produce deterministic routes and failures during autonomous driving. A custom Python playback application rendered the recorded frames of the vehicle front camera, applied PSAE overlays frame-by-frame, and synchronized playback with participants' steering and braking inputs; steering angle and pedal signals were logged over Ethernet for offline alignment.

Physiological measures comprised continuous EEG recorded with a Neuroelectrics Enobio 32-channel dry sensors[1] at 500 Hz, and surface EMG recorded at 1000 Hz through PhysioLAB platform[2] from two sites: the dominant-hand forearm (approximately the flexor region, to index steering-related activation) and the right medial gastrocnemius (to index braking/leg activation) [81, 90]. All data streams (simulator telemetry, AR-HUD rendering timestamps, EEG, EMG, and behavioral logs) were synchronized via millisecond timestamps to enable precise multimodal analyses.

3.5 Data Processing and Variable Extraction

EEG. The preprocess of EEG signal followed a standardized MNE pipeline [38] to ensure high signal quality and reproducibility: band-pass filtering between 1–30 Hz was applied to remove slow drifts and high-frequency noise, common average re-referencing, and Independent Components Analysis (ICA) was used to isolate and remove components associated with eye blinks and ocular movements [25]. Channels with excessive noise (variance > 3 standard deviations from the mean) were interpolated using spherical spline interpolation [76]. To quantify neural dynamics related to experimental events, we computed the relative ERSP change ratio (*Power Ratio*), which measures changes in rhythmic brain activity power relative to the pre-event baseline period.

Specifically, for each 10-second event, we extracted a 4-second pre-event baseline from the raw EEG time series. The baseline series was converted to a time-frequency representation via Morlet wavelets, which provides an optimal trade-off between temporal and frequency resolution for analyzing cognitive processes [14]. For each scene's event and baseline power, we computed the median power per channel and per frequency across time. The resulting power at each time point ($P_{event}(t)$) at every event window was normalized against the median power from a 4-second pre-event baseline period ($P_{baseline}$) using the formula:

$$\text{ERSP}(t) = \frac{P_{event}(t) - P_{baseline}}{P_{baseline}}. \tag{1}$$

We summarized ERSP within canonical frequency bands (i.e., θ (4–8 Hz), α (8–12 Hz), and β (15–30 Hz)) and averaged these across three large-scale Region-of-Interests (ROIs) (i.e., Frontal,

Sensorimotor, Parietal) to capture dynamics related to attention, motor planning, and sensory integration [60]. From each ROI–band time–frequency matrix, we extracted two features: (1) the median *Power Ratio* across the event windows time, and (2) *Activation Time*, the latencies of the peak responses (for Beta the time of maximal synchronization; for Theta/Alpha the time of maximal desynchronization [77]), which together characterize both magnitude and temporal dynamics of neural responses.

Table 2: Summary of dependent variables across different layers.

Layer	Variable [unit]	Distribution
Psychological Perception	Situation Awareness [-]	Mean: 5.2 (SD: 0.9, min: 2.3, max: 7)
	Perceived Safety [-]	Mean: 4.1 (SD: 1.0, min: 1.0, max: 6)
	Trust [-]	Mean: 45.7 (SD: 27.1, min: 0, max: 91)
Cognitive Activity	S_β Activation Time [ms]	Mean: 3613 (SD: 3519, min: 2, max: 10000)
	F_θ Activation Time [ms]	Mean: 5510 (SD: 2633, min: 2, max: 10000)
	P_α Activation Time [ms]	Mean: 5192 (SD: 2669, min: 112, max: 9828)
	S_β Power Ratio [-]	Mean: 0.3 (SD: 2.2, min: -0.9, max: 5.5)
	F_θ Power Ratio [-]	Mean: 1.05 (SD: 10.4, min: -1, max: 23.7)
	P_α Power Ratio [-]	Mean: 0.3 (SD: 2.3, min: -1, max: 29.3)
Muscle Preparation	Preparation [-]	No preparation (n=28, 12.8%) Preparation exist (n=190, 87.2%)
	Preparation Time [ms]	Mean: 3347 (SD: 3885, min: 0, max: 9999)
Brain–Muscle Coupling	S_β Time Lag [ms]	Mean: 266.3 (SD: 4481, min: -9977, max: 10000)
	F_θ Time Lag [ms]	Mean: 2163 (SD: 4687, min: -9378, max: 10000)
	P_α Time Lag [ms]	Mean: 1845 (SD: 4588, min: -9862, max: 9120)
Performance	Takeover Lead Time [ms]	Mean: 1106 (SD: 982, min: 0, max: 5800)
	Takeover Success [-]	No takeover (n=32, 14.5%) ∼ 0 Success takeover (n=186, 85.5%) ∼ 1

Notes: S_β: Sensorimotor Beta; F_θ: Frontal Theta; P_α: Parietal Alpha; SD: standard deviation

EMG. To quantify participants' motor preparation, we processed surface EMG signals recorded from the hand and leg. The raw signals first underwent preprocessing, which included a zero-phase fifth-order Butterworth band-pass filter to isolate the primary EMG frequency content [23], followed by full-wave rectification (i.e., taking the absolute value) and normalization to prepare the signal for amplitude analysis [78]. We then identified muscle activation onset by detecting the first sample in the rectified signal that exceeded a predefined activation threshold 0.1 (i.e., 10% of the normalized EMG amplitude [5]) for EMG-based onset detection [42]. This onset time, calculated in milliseconds relative to the end of the trial window, served as our primary measure of motor preparation latency. Furthermore, based on the binary activation status of the hand and leg channels, we classified each trial into preparation exists and no preparation (*Preparation*), and the overall *Preparation Time* was defined as the earliest muscle activation onset across the measured limbs.

Brain–Muscle Coupling. To characterize the temporal coordination between cortical processes and peripheral muscle execution, we computed the latency difference (*Time Lag*) between the peak of the ERSP in the EEG and the onset of muscle activation in the EMG:

$$\textit{Time Lag} = t_{\text{EEG\_peak}} - t_{\text{EMG\_onset}}. \tag{2}$$

---

[1]https://www.neuroelectrics.com/products/research/enobio/enobio-32
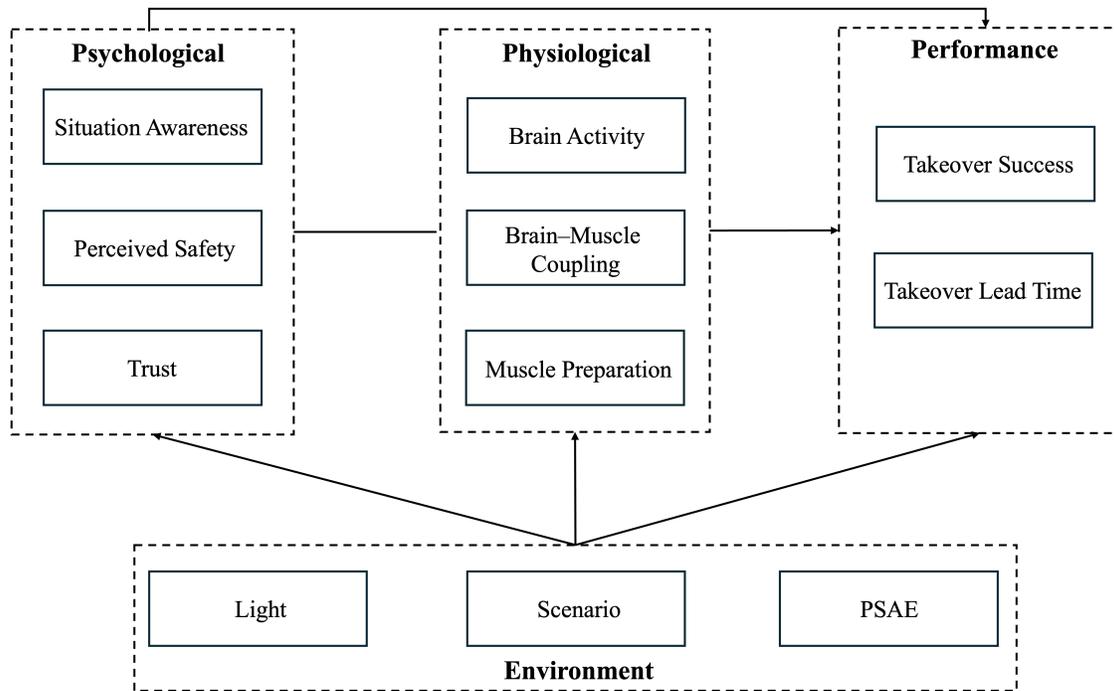[2]https://www.infoinstruments.cn/product/physiolab/

Figure 4: The final theoretical framework of the model.

While the onset of motor-related potentials (e.g., the Readiness Potential) of EEG offers a more direct precursor to movement, the ERSP peak was chosen for two reasons. First, it represents the point of maximal cortical disinhibition and engagement in motor planning for this reactive task, capturing the climax of the decision-to-act process [77]. Second, peak detection is more robust to the signal-to-noise limitations inherent in dry-EEG recordings compared to threshold-based onset detection. Therefore, our latency measure characterizes the timing between peak cortical activity and peripheral execution. A summary of all extracted variables and descriptive statistics is provided in Table 2.

3.6 Statistical Analysis

Based on previous theory frameworks [32, 66] and previous empirical studies [18, 30, 92, 99, 102], our analytical approach is guided by a theoretical framework that decomposes the driver's response into a multi-layered causal sequence (see Figure 4). Our analytical approach was designed to systematically test the relationships within our multi-layered theoretical framework, which posits that the *Environment* layer influences *Psychological* layer. Then *Psychological* layer is further associated with *Physiological* layer, and they jointly or independently affect the final takeover *Performance* layer.

The analysis proceeded in two main stages: preliminary regression analyses, followed by a formal path analysis to test for mediation. First, to establish the foundational relationships between the layers, we conducted a series of regression analyses using SAS OnDemand for Academics. The model type was chosen based on the outcome variable: Generalized Linear Models (via PROC GENMOD) for binary outcomes (i.e., takeover success), and Linear Mixed-effects Models (via PROC MIXED) for continuous variables such as takeover time. To properly handle the non-independence

of repeated measures, all mixed models incorporated a random intercept for each participant, thereby accounting for individual differences. Initial models included the main effects of our experimental conditions and variables from the preceding layers (see Figure 4).

At the same time, we are interested in the moderating effects of PSAE in this process. Therefore, when constructing the models for the variables in the *Physiological* and *Performance* layers, we incorporated two-way interaction between SA and PSAEs as independent variables. A backward stepwise selection procedure was then used to arrive at the most parsimonious model, with multicollinearity confirmed to be absent via the Variance Inflation Factor (VIF). Following significant main or interaction effects ($p<.05$), Tukey-Kramer post-hoc tests [52] were performed. Note that, before fitting the models, we performed power checks based on our experiment design. Following [12], we considered medium (f=0.25) and large (f=0.40) effects. Results suggested that our sample size affords high power to detect large main effects of PSAE and lighting (power>0.85 for f=0.40), but only moderate power for medium effects (power=0.43 for PSAE and power=0.61 for lighting condition) and limited power for interactions across multiple experimental factors. To avoid underpowered and unstable estimates under these constraints, we therefore (a) focused on main effects in the mixed-effects and generalized linear models, and (b) restricted interaction terms to a small set of two-way interactions between experimental factors that showed significant main effects and continuous psychological or physiological measures (e.g., SA × PSAE), without specifying interactions among experimental factors.

Finally, to further explore how the PSAE affects the takeover performance, we conducted a path analysis using Structural Equation Modeling (SEM) via the lavaan package in R [80]. To account

Table 3: Summary of statistical results.

Dependent Variable (DV)	Independent Variable (IV)	F-value/χ^2-value	Estimate (95% CI)	p	Fit indicator
Situation Awareness	Light	F(1, 40) = 4.69	–	.03*	Adj R^2 = 0.08
	PSAE	F(3, 40) = 5.11	–	.002*	p = .001*
Perceived Safety	PSAE	F(3, 42) = 1.85	n.s.	.1	Adj R^2 = 0.15
	Light	F(1, 40) = 1.00	n.s.	.2	p < .0001*
	Scenario	F(1, 40) = 30.81	–	<.0001*	
Trust	PSAE	F(3, 40) = 6.43	–	.0003*	Adj R^2 = 0.12
	Scenario	F(1, 42) = 5.57	–	.02*	p < .0001*
	Light	F(3, 40) = 4.61	–	.03*	
Preparation Time	PSAE	F(3, 42) = 2.69	–	.048*	
	Light	F(1, 40) = 1.01	n.s.	.3	Adj R^2 = 0.07
	Scenario	F(1, 40) = 0.44	n.s.	.5	p = .03*
	Trust	F(1, 238) = 0.72	n.s.	.4	
	Situation Awareness*PSAE	F(3, 238) = 1.86	n.s.	.1	
Preparation	Light	χ^2(1) = 6.64	–	.01*	pseudo-R^2 = 0.09 p = .001*
F_θ Activation Time	PSAE	F(3, 40) = 3.36	–	.02*	
	Scenario	F(1, 40) = 4.08	–	.045*	Adj R^2 = 0.11
	Situation Awareness	F(1, 238) = 2.72	n.s.	.053	p = .01*
	Situation Awareness*PSAE	F(3, 238) = 2.48	–	.0498*	
P_α Activation Time	Situation Awareness	F(1, 239) = 4.41	431.42 [25.17, 837.66]	.04*	
	Perceived Safety	F(1, 239) = 0.39	n.s.	.5	Adj R^2 = 0.04
	PSAE	F(3, 42) = 1.07	n.s.	.4	p = .02*
	Situation Awareness*PSAE	F(3, 239) = 0.83	n.s.	.5	
F_θ Power Ratio	PSAE	F(3, 42) = 4.12	–	.001*	
	Trust	F(1, 238) = 1.83	n.s.	.2	Adj R^2 = 0.07
	Situation Awareness	F(1, 238) = 3.91	n.s.	.0501	p = .045*
	Situation Awareness*PSAE	F(3, 238) = 3.21	–	.03*	
P_α Power Ratio	PSAE	F(3, 42) = 2.01	n.s.	.2	Adj R^2 = 0.05
	Situation Awareness	F(1, 238) = 7.31	-0.48 [-0.83, -0.13]	.008*	p = .03*
	Trust	F(1, 238) = 4.03	-0.01 [-0.02, -0.0002]	.047*	
P_α Time Lag	Situation Awareness	F(1, 239) = 4.98	907.03 [103.75, 1710.31]	.03*	Adj R^2 = 0.03 p = .02*
Takeover Success	P_α Activation Time	χ^2(1) = 4.93	0.0002 [0.0, 0.0004]	.03*	
	P_α Time Lag	χ^2(1) = 2.19	n.s.	.1	
	Light	χ^2(1) = 29.00	–	<.0001*	
	PSAE	χ^2(3) = 4.88	n.s.	.2	pseudo-R^2 = 0.48
	Scenario	χ^2(1) = 19.43	–	<.0001*	p < .0001*
	Situation Awareness	χ^2(1) = 8.32	1.64 [0.47, 2.82]	.004*	
	Perceived Safety	χ^2(1) = 8.33	0.96 [0.31, 1.61]	.004*	
	Situation Awareness*PSAE	χ^2(3) = 8.93	–	.03*	
Takeover Lead Time	Light	F(1, 40) = 15.91	–	.0003*	Adj R^2 = 0.36
	Scenario	F(1, 40) = 8.08	–	<.0001*	p < .0001*
	Situation Awareness	F(1, 235) = 4.85	157.11 [16.03, 298.19]	.03*	

Notes: In this table and the following tables, * marks significant results (p<.05). In the Estimate (95% CI) column, "n.s." means nonsignificant (p > 0.05), and "–" means the dependent variable is a nominal variable so that there are no Estimate values. The "Fit indicator" column reports, for each dependent variable, the adjusted R^2 (for linear mixed-effects models) or a McFadden-type pseudo-R^2 (for generalized linear models), and p-value when comparing the model to the null model without predictors.

for repeated measurements, we adopted the Mixed-Effects Model (MLM) in SEM to account for individual and experimental sequence effects [92]. This approach avoids the violation of the SEM assumptions on data independence. To compare the effects of each PSAE condition against the baseline, we created dummy variables for the *PSAE* factor. For the continuous variable, we used a robust maximum likelihood estimator (MLR) and tested the significance of the indirect effect using bias-corrected confidence intervals from 5,000 bootstrap resamples. For the binary variable, we declared it

as an ordered factor and used the robust weighted least squares estimator (WLSMV), which is appropriate for categorical outcomes. For both models, our primary goal was to quantify and test the significance of the specific indirect effect for each PSAE condition, thereby revealing the underlying mechanism through which the interface influenced driver performance.

Table 4: Significant Post-hoc Results for Discrete Independent Variables.

DV	IV	IV Level	IV Level compared to	Estimation (95% CI)	t value/z score	p
Situation Awareness	Light	Night	Day	-0.25 [-0.48, -0.02]	t(40) = -2.17	.03*
	PSAE	EP	EP+PM	0.45 [0.04, 0.87]	t(42) = 2.86	.02*
		EP	Baseline	0.61 [0.17, 1.04]	t(42) = -2.17	.002*
Perceived Safety	Scenario	Visible	Invisible	0.51 [0.33, 0.70]	t(40) = 5.55	<.0001*
Trust	Light	Night	Day	-7.57 [-14.53, -0.62]	t(40) = -2.15	.03*
	PSAE	PM	Baseline	15.52 [2.35, 28.69]	t(42) = 3.05	.01*
		EP+PM	Baseline	20.76 [8.01, 33.50]	t(42) = 4.22	<.0001*
	Scenario	Visible	Invisible	8.23 [1.36, 15.12]	t(40) = 2.36	.02*
Preparation	Light	Night	Day	1.42 [0.34, 2.50]	z = 2.58	.01*
F_θ Activation Time	Scenario	Visible	Invisible	765.35 [16.16, 1514.54]	t(20) = 2.02	.045*
Takeover Success	Light	Night	Day	-2.70 [-3.69, -1.72]	z = -5.39	<.0001*
	Scenario	Visible	Invisible	2.23 [1.24, 3.22]	z = 4.41	<.0001*
Takeover Lead Time	Light	Night	Day	-587.36 [-885.44, -289.28]	z = -3.99	.0003*
	Scenario	Visible	Invisible	1056.78 [828.42, 1285.14]	z = 9.39	<.0001*

Notes: Non-significant variables are not listed in this table.
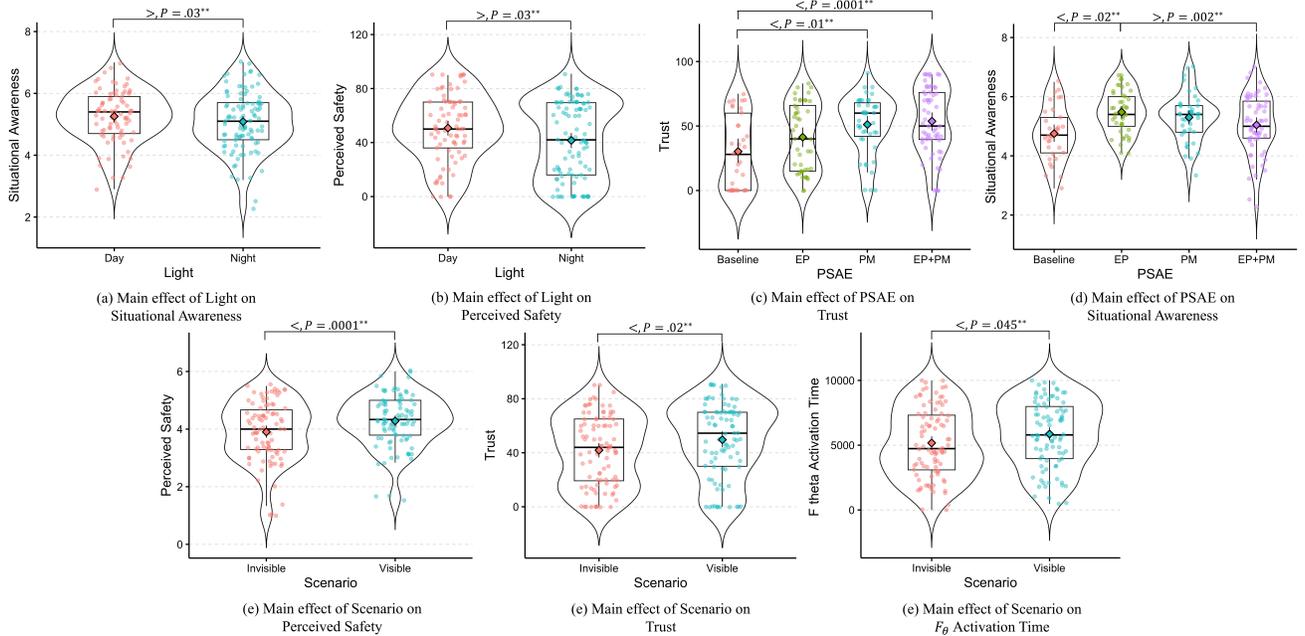
Figure 5: The main effects of experimental conditions on drivers' psychological and physiological states. (a) The effect of the time of day on drivers' situation awareness. (b) The effect of the time of the day on perceived safety. (c) The effect of PSAE information on trust. (d) The effect of PSAE information on situation awareness. (e) The effect of scenario type on perceived safety. (f) The effect of scenario type on trust. (g) The effect of scenario type on F_θ Activation Time.

4 Results

4.1 Regression Analysis

This section presents the results from regression analyses, which were conducted as the first stage of our analytical approach to establish the foundational relationships within our theoretical framework (RQ1 and RQ2). The statistical summaries are presented in Table 3, and the significant post-hoc comparisons are detailed in Table 4. In addition to the fixed-effect statistics, we also report overall model fit indicators. For continuous outcomes, we fit an equivalent fixed-effects linear model and report the resulting adjusted

coefficient of determination (adjusted R^2) and p-value from the overall model F-test as descriptive indices of how much variance is accounted for by the fixed effects [13]. For binomial outcomes, we report a McFadden-type pseudo-R^2 [63] and p-value in chi-square model test when comparing the current model versus intercept-only model. Visualizations of main effects and interaction effects are provided in Figure 5 and 6, respectively. Note that, given the number of models and predictors considered, and the limited sample size in this study, the regression results in Table 3 should be viewed as
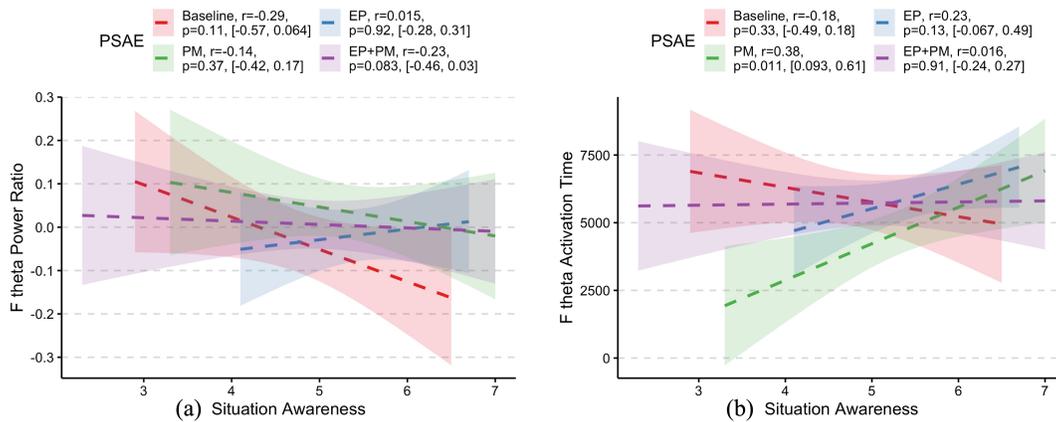
Figure 6: Interaction effects between the Situational Awareness and PSAE on (a) F_θ Power Ratio and (b) F_θ Activation Time. The samples of each PSAE group are fitted by a regression line with a 95% confidence interval. Legends contain the results of correlation analysis based on Spearman's correlation.

exploratory. The key paths suggested by these exploratory regressions are subsequently evaluated in a more integrated manner in the SEM.

In response to RQ1, our regression results revealed that both the PSAE interface and lighting conditions were significant predictors of drivers' psychological states. Specifically, *PSAE* had a significant main effect on both SA, which was measured using the SART scale, and Trust. The PM and EP+PM interfaces significantly increased driver *Trust* compared to the Baseline condition. For *Situation Awareness*, the results were more complex; the EP interface was associated with significantly higher *Situation Awareness* compared to both the Baseline and the integrated EP+PM interface. The *Light* condition also showed significant main effects on SA and Trust. Interestingly, drivers reported significantly lower *Situation Awareness* during night-time drives compared to daytime, while simultaneously reporting lower *Trust* in the system. Finally, the *Scenario* type, as expected, was a powerful predictor for Perceived Safety and Trust. Environmental factors also influenced physiological readiness. The *Light* condition had a significant effect on muscle *Preparation*, with post-hoc tests showing a higher likelihood of preparation during night-time, corroborating the vigilance effect observed in the psychological data. The *PSAE* interface also demonstrated a significant main effect on *Preparation Time* and F_θ *Power Ratio*. More importantly, we found significant interaction effects between *PSAE* and *SA* on several physiological variables, including F_θ Activation Time and F_θ Power Ratio, suggesting that the influence of the PSAE on drivers' neural activity is contingent upon their concurrent psychological state.

In response to RQ2, the results indicated various relationships between psychological states and neural activity. As shown in Table 3, higher *Situation Awareness* was a significant predictor of decreased P_α Power Ratio, as well as earlier P_α Activation Time and a larger P_α Time Lag. Higher SA was also significantly associated with an increase in F_θ Power Ratio. Furthermore, higher *Trust* was also linked to a decrease in P_α Power Ratio.

As a foundational step for our subsequent path analysis (RQ3), we found that *Situation Awareness* was a strong predictor of takeover

performance. Higher SA was significantly associated with a higher likelihood of a *Takeover Success* and an earlier *Takeover Lead Time*.

4.2 Path Analysis

Based on our preliminary regression results, which identified SA as a critical hub, we conducted a formal path analysis to test the mediation mechanism proposed in RQ3. We specified a simple mediation model (*PSAE -> SA -> Performance*). The measurement model for the SA latent variable demonstrated good internal consistency, with all factor loadings exceeding a threshold of 0.5. The path models for both performance metrics achieved a good overall fit [92].

As shown in Table 5, the path analysis provided clear evidence that SA acts as a key mediator for the effect of the PSAE interface on takeover success. We found significant indirect effects via SA for all three PSAE conditions when compared to the baseline: for EP (0.431, 95% CI [0.173, 0.688], $p = <.0001^*$), PM (0.327, 95% CI [0.084, 0.570], $p = .008^*$), and EP+PM (0.209, 95% CI [0.019, 0.400], $p = .03^*$). At the same time, the direct effects of the PSAE conditions on Takeover Success (i.e., c' paths) were small and nonsignificant. These results are therefore statistically consistent with full mediation, meaning that PSAE increased takeover success primarily through improvements in drivers' SA rather than via a direct effect on behavior.

For Takeover Lead Time, the EP condition showed a significant indirect effect via SA (indirect = 0.170, 95% CI [0.018, 0.321], $p = .03^*$). The PM condition demonstrated a borderline indirect effect that approached significance (indirect = 0.135, 95% CI [0.000, 0.270], $p = .051$), and EP+PM did not show a significant indirect effect. Here, we therefore describe PM as showing a trend-level mediation effect and EP as showing a robust mediation effect.

Furthermore, we conducted a series of path analyses to test the alternative hypothesis that Situation Awareness might also mediate the effect of PSAE on key physiological indicators. However, none of these models achieved an acceptable level of model fit (e.g., all CFIs < 0.89, all RMSEAs > 0.12), indicating a significant discrepancy between our hypothesized causal structure and the observed data. As the models themselves were misspecified, the path coefficients

Table 5: Path analysis results (a / b / c' / indirect).

DV	Path Type	Path	Estimation [95% CI]	Std. β	p
	a	EP → SA	0.840 [0.369, 1.312]	0.382	<.0001*
	a	PM → SA	0.638 [0.196, 1.079]	0.285	.005*
	a	EP+PM → SA	0.409 [0.043, 0.775]	0.201	.03*
	b	SA → Takeover Success	0.512 [0.330, 0.694]	0.482	<.0001*
Takeover Success	c'	EP → Takeover Success	0.009 [-0.679, 0.696]	0.004	.98
	c'	PM → Takeover Success	-0.342 [-0.998, 0.313]	-0.144	.3
	c'	EP+PM → Takeover Success	-0.018 [-0.636, 0.601]	-0.008	.96
	indirect (a×b)	EP → SA → Takeover Success	0.431 [0.173, 0.688]	0.184	<.0001*
	indirect (a×b)	PM → SA → Takeover Success	0.327 [0.084, 0.570]	0.137	.008*
	indirect (a×b)	EP+PM → SA → Takeover Success	0.209 [0.019, 0.400]	0.097	.03*
	a	EP → SA	0.805 [0.380, 1.231]	0.354	<.0001*
	a	PM → SA	0.640 [0.193, 1.087]	0.277	.005*
	a	EP+PM → SA	0.416 [-0.072, 0.903]	0.198	.095
	b	SA → Takeover Lead Time	0.211 [0.048, 0.374]	0.186	<.0001*
Takeover Lead Time	c'	EP → Takeover Lead Time	0.129 [-0.323, 0.581]	0.050	.6
	c'	PM → Takeover Lead Time	0.141 [-0.330, 0.613]	0.054	.6
	c'	EP+PM → Takeover Lead Time	0.361 [-0.067, 0.789]	0.152	.098
	indirect (a×b)	EP → SA → Takeover Lead Time	0.170 [0.018, 0.321]	0.066	.03*
	indirect (a×b)	PM → SA → Takeover Lead Time	0.135 [0.000, 0.270]	0.051	.051
	indirect (a×b)	EP+PM → SA → Takeover Lead Time	0.088 [-0.027, 0.203]	0.037	.1

Notes: In Path Type, "a" denotes the path from PSAE condition to SA; "b" denotes the path from SA to the DV; "c' " denotes the direct effect of the PSAE condition on the DV controlling for SA; "indirect" denotes the a×b mediated effect. "Estimation" reports unstandardized coefficients with bias-corrected bootstrap 95% CI (N = 5,000 resamples). "Std. β" reports standardized coefficients. Indirect-effect significance was assessed via bootstrap CI (significant if 95% CI does not include 0).

within them cannot be reliably interpreted. This lack of fit suggests that a simple PSAE -> SA -> Physiological causal chain may not be appropriate for our data. The detailed fit indices and parameter estimates for these exploratory but ultimately rejected models are available in the Appendix Table 6.

5 Discussion

5.1 RQ1: How Interfaces and Environment Shape Driver States

Our first research question explored the direct impact of the PSAE interfaces and environmental conditions (lighting and hazard visibility) on drivers' SA, perceived safety, trust, and physiological indicators (EEG/ERSP and EMG measures). The findings confirm that these external factors are influential modulators of drivers' cognitive activities and motor preparation, particularly within the context of a silent failure. A key finding is that the design of the PSAE interface significantly impacts drivers' trust and perceived safety. The PM and EP+PM interfaces notably increased driver trust, suggesting that providing information about the system's planned maneuvers is crucial for building confidence [58].

While abundant information can sometimes cause overload, in our data, the EP interface (displaying environment perception) was associated with significantly higher Situation Awareness compared to Baseline and EP+PM. This suggests that presenting perceptual cues can help drivers better perceive the scene in silent-failure situations. At the same time, the comparatively lower SA in mixed information presentation (EP+PM) suggests that excessive or poorly organized visual information (including AR/highlight displays and

dense HUD content) might increase workload and impair SA in driving contexts [35].

Furthermore, we uncovered a "night-time vigilance effect." Drivers in night-time conditions reported significantly higher SA but lower Trust, a finding corroborated by a higher likelihood of muscle preparation. This suggests a compensatory mechanism: when faced with a silent failure under inherently riskier conditions, drivers' reduced trust may compel them to maintain a higher state of vigilance, which aligns with literature on vigilance fluctuations and visibility-related effects on driving performance and preparedness [45]. This highlights the impact of lighting on drivers' SA, trust, and motor preparation when handling hazards. Finally, the discovery of a significant interaction effect between PSAE and SA on neural activity (e.g., F_θ Power Ratio) suggests that the way an interface supports drivers' cognitive processing during a potential silent failure is not uniform, but is contingent on the driver's SA level (as measured by SART) [6].

5.2 RQ2: Uncovering the Link Between Mind and Body

Our second research question sought to delineate the relationship between subjective psychological states (SA, perceived safety, trust) and objective physiological indicators derived from EEG and EMG. The most compelling finding here is the consistent and strong association between SA and Parietal Alpha (P_α) activity. Our results clearly showed that higher SA was linked to a lower P_α Power Ratio (i.e., alpha suppression) and earlier P_α activation time. This aligns with established cognitive neuroscience literature, where the suppression of parietal alpha power is a well-established neural

marker of heightened visuospatial attention and cognitive engagement [26, 103]. This convergence validity between subjective and objective measurements has, to a certain extent, strengthened our confidence in using SART scores as a proxy indicator for SA. Meanwhile, our findings empirically ground this theoretical link in a dynamic driving task. This is critical in the context of silent failures, as this neural activity might represent the very first moment the driver's brain flags a problem that the automation might have encountered. This suggests P_α activity can serve as a potential neural correlate of the implicit subjective SA state.

Interestingly, while SA was strongly linked to cortical activity (EEG), its direct link to peripheral motor readiness (*Preparation Time*) was not significant. Meanwhile, we found that higher SA was associated with a larger EEG–EMG time lag. This "disconnect" suggests that, while cortical engagement (as indexed by an earlier alpha modulation) marks faster detection/processing of the scene, the increased EEG–EMG lag suggests that this cortical detection does not always translate immediately into peripheral motor commitment. In other words, cognitive detection and motor initiation can be decoupled in silent-failure contexts, which was formally tested in our path analysis. Such a phenomenon echoes previous findings that a good SA does not always mean a good decision or action [71].

We also observed that higher SA tended to go together with increased frontal/theta-band engagement (F_θ Power). Frontal midline theta has been widely interpreted as reflecting cognitive control, conflict monitoring, and the mobilization of executive resources, which are expected when a driver detects a mismatch between the system's behavior and the environment [6]. Thus, the joint pattern of parietal alpha suppression and frontal theta increase is neurophysiologically coherent: the better SA is associated with enhanced sensory/attentional processing (especially in the case of ADS failures), which facilitates control/monitoring processes [89].

5.3 RQ3: The Mediating Role of SART in Driving Performance

Our third research question investigated the underlying mechanism through which PSAE interfaces affect performance. The path analysis results provide evidence that the effect of the PSAE interface on takeover performance is, to a large extent, indirectly associated with Situation Awareness. Within our experiment, this mediation appears to be an important mechanism for supporting drivers when handling a silent failure: the interface's primary role may be to facilitate the driver's self-detection of the problem (by enhancing SA) when the system itself provides no alarm. Prior work on agent transparency and attention-guiding takeover requests similarly suggests that interface transparency and attention guidance improve SA and thereby support takeover performance [8, 91] in TOR-initiated takeover scenarios.

Our models for both Takeover Success and Takeover Lead Time converged on this same conclusion. The pattern in the results, with significant indirect effects but without significant direct effects, is consistent with a model in which SA serves as a central pathway through which the interface can influence performance. It is noteworthy that we tested an alternative hypothesis, that SA might mediate the effect of PSAE on physiological states, but found no supporting evidence. This crucial null finding suggests that while physiological states are correlated with driver cognition (as in RQ2), they do not appear to be part of the primary causal chain that flows from the HMI, through SA, to performance in our current models.

This null mediation may reflect several possibilities: a mismatch in the timescales between subjective SA reports and millisecond-level EEG changes, limitations in the signal-to-noise ratio of the physiological measures, or that the true relationship is more complex, perhaps involving conditional effects (i.e., moderated mediation) that our current sample size lacks the power to detect. Future work should use measures with higher temporal resolution to test for potential conditional mediation pathways.

5.4 Implications

The collective findings of this study, grounded in the challenging context of silent automation failures, offer several important implications. Our results are consistent with and further support the view that SA is an important construct for safe human-automation interaction [32]. The mediation pattern observed in response to RQ3 offers one possible explanation in which a key function of transparency-oriented HMIs in silent failures is to enhance the driver's cognitive understanding of the scenario, which in turn enables effective action. This interpretation is also aligned with the hierarchical model of driving, where tactical-level SA is a prerequisite for operational control [66], and recent theories that treat takeover as a decision process driven by evidence accumulation and risk appraisal [36, 62, 88].

The consistent relationship between SA and parietal alpha suppression has both theoretical and applied implications. Theoretically, it anchors the abstract construct of SA to an empirically tractable neurophysiological marker (posterior alpha dynamics), strengthening the bridge between neuroergonomics and HMI design. Practically, the consistent relationship between SA and parietal alpha suppression raises the possibility of augmenting HMI systems with neuro-informed monitoring. If further research can confirm that the parietal alpha suppression can reliably index SA in real driving contexts, EEG-derived metrics could potentially be used to support adaptive HMIs that aim at increasing SA. However, three important caveats apply. First, scalp alpha measures can be noisy in real-world settings and require robust preprocessing and individualized baselining [26, 89]. Second, the temporal relationship we observed between cortical markers and peripheral motor preparation is complex, implying that neurofeedback alone may not be sufficient to predict imminent action. Third, any application that uses neural measures to adapt HMI must be validated for generalizability across drivers, contexts, and device types [94].

Our findings also provide guidance for the design of in-vehicle interfaces. The results related to RQ1 suggest that more information is not always better. Presenting perceptual information in a clear, salient, and non-competing manner is crucial: poorly organized HMIs may impede drivers' comprehension of the situation. Future HMI designs should not just focus on the type of information to display, but also how to integrate and present the information effectively. For example, we even found a reduced SA under the combined 'EP+PM' interface. This serves as a critical cautionary

tale for HMI design: simply adding more information, even if theoretically useful, can lead to cognitive overload or visual interference, ultimately harming performance. Finally, the discovery of the "night-time vigilance effect" highlights that the effectiveness of HMI might be dynamic. This motivates a shift away from "one-size-fits-all" interfaces towards adaptive systems that can tailor information delivery to the specific situation and driver [27]. The relationship between physiological and cognitive states and behavior also informs the future design of adaptive HMIs. For example, in-vehicle driver monitoring systems [95, 98] can obtain real-time information on the driver's physiological [96, 100] and cognitive states [93, 101] and then flexibly adjust the HMI according to the in- or out-of-cabin context.

6 Limitations

We recognize that our study, while following rigorous methods, has several limitations that offer avenues for future research. First, our between-subjects design for the PSAE factor, while necessary to avoid carryover effects, resulted in a small sample size per group. Besides, our regression models show modest but acceptable R^2 [11, 70]. We therefore interpret these associations as statistically reliable but modest in size, and do not claim high predictive accuracy at the individual level. These limit the statistical power to detect smaller or more complex interaction effects and necessitate caution when generalizing our findings. Future work could benefit from larger samples or exploring alternative designs. Second, while our high-fidelity driving simulator provided a safe and controllable environment for studying hazardous events, it cannot fully replicate the physical and emotional risks of real-world driving. The psychological and physiological responses observed may differ from those in on-road situations, and future studies should seek to validate our findings in more naturalistic settings.

Further, our path analysis revealed that the proposed physiological mediators have a poor overall fit. While this suggests such simple mediation is not the correct mechanism, it may also reflect several possibilities: a mismatch in the timescales between subjective SA reports and millisecond-level EEG changes, limitations in the signal-to-noise ratio of the physiological measures, or that the true relationship is more complex, perhaps involving conditional effects (i.e., moderated mediation) that our current sample size lacks the power to detect. Moreover, we operationalized takeover performance at the level of immediate response to hazard, i.e., how much lead time the driver had when taking control, without analysing post-takeover driving quality (e.g., lane keeping, speed, or headway control after the intervention). Future work could examine how post-takeover driving performance can be influenced by the visualization of ADS states, which may be affected by different factors through different pathways. Next, our study focused on a specific set of silent failure types and environmental factors. The effectiveness of these PSAE interfaces may vary across different types of hazards, traffic densities, or weather conditions. Other forms of silent failures, such as control-level faults and sudden system shutdowns, were not considered in this study. Future research should explore a wider range of scenarios and silent failure types to build a more comprehensive understanding of how to support drivers in more diverse silent failure scenarios. Finally, we used

questionnaire results (e.g., SART, trust and perceived safety) as subjective measures of perceived psychological states. Although we only modeled them as indirect indicators of cognitive processes, it is important to recognize that the self-reported questionnaires reflect participants' retrospective and overall impressions of each scenario. Such ratings are often influenced by factors like knowledge of the outcome, recall bias, and self-presentation tendencies, and do not offer a moment-to-moment and objective account of drivers' psychological states during the event. Therefore, the readers should be cautious when interpreting our model and future research may consider objective measures of SA, trust and perceived safety.

7 Conclusion

In this study, we conducted a multi-modal driving simulator experiment involving 48 participants to investigate whether different Prospective Situation Awareness Enhancement (PSAE) interfaces can support drivers during safety-critical silent automation failures. By integrating psychological surveys, neurophysiological recordings (EEG, EMG), and behavioral data, we systematically analyzed the pathways from interface design to driver performance. Our analyses, combining mixed-effects regression and path analysis, yielded several key findings:

- In our study, Situation Awareness (SA) emerged as a key pathway through which PSAE interfaces were associated with takeover performance. Path analysis revealed a significant indirect effect where the interfaces first influenced SA, which in turn predicted both takeover success and lead time. This highlights that the principal goal for designers of transparency-focused HMIs should be to maximize driver comprehension of the scenario (including both ADS and the traffic environment).
- Different PSAE interfaces had different effects: EP (perception information) improved SA, while interfaces including planned maneuver information (PM and EP+PM) were only effective in increasing Trust. These results suggest that future in-vehicle HMI designs should consider balancing the richness of the information to increase users' trust in the system while avoiding overloading drivers.
- We observed patterns consistent with a potential neural correlate of SA. Higher SA scores were associated with suppression of alpha-band activity, earlier neural activation time, and a longer time lag between neural and motor responses. These findings can motivate further work on physiological-informed driver-state monitoring [95, 98] and contextualized and adaptive HMIs, though rigorous field validation with more diverse individuals is still needed.
- Finally, we observed that the effectiveness of HMI in our study appeared to be context-dependent. The "night-time vigilance effect," where drivers exhibited higher SA but lower trust at night, suggests that external conditions can influence a driver's internal state. Together with the interaction effects between PSAE and SA, this finding indicates the need for exploring context-aware and adaptive systems that can tailor information delivery to specific situations and heterogeneous drivers.

Acknowledgments

References

[1] Mohanad Abukmeil, Angelo Genovese, Vincenzo Piuri, Francesco Rundo, and Fabio Scotti. 2021. Towards explainable semantic segmentation for autonomous driving systems by multi-scale variational attention. In *2021 IEEE International Conference on Autonomous Systems (ICAS)*. IEEE, 1–5.

[2] Giulio Bianchi Piccinini, Esko Lehtonen, Fabio Forcolin, Johan Engström, Deike Albers, Gustav Markkula, Johan Lodin, and Jesper Sandin. 2020. How do drivers respond to silent automation failures? Driving simulator study and comparison of computational driver braking models. *Human factors* 62, 7 (2020), 1212–1229.

[3] National Transportation Safety Board. 2018. Preliminary Report: Highway (HWY18MH010). *Preliminary Report on First Autonomous Vehicle Test Car Fatality in Arizona on* (2018).

[4] Jianqin Cao, Li Lin, Jingyu Zhang, Liang Zhang, Ya Wang, and Jifang Wang. 2021. The development and validation of the perceived safety of intelligent connected vehicles scale. *Accident Analysis & Prevention* 154 (2021), 106092.

[5] Camila R Carvalho, J Marvin Fernández, Antonio J Del-Ama, Filipe Oliveira Barroso, and Juan C Moreno. 2023. Review of electromyography onset detection methods for real-time control of robotic exoskeletons. *Journal of neuroengineering and rehabilitation* 20, 1 (2023), 141.

[6] James F Cavanagh and Michael J Frank. 2014. Frontal theta as a mechanism for cognitive control. *Trends in cognitive sciences* 18, 8 (2014), 414–421.

[7] Kuan-Ting Chen, Huei-Yen Winnie Chen, Ann Bisantz, Su Shen, and Ercan Sahin. 2023. Where failures may occur in automated driving: a fault tree analysis approach. *Journal of cognitive engineering and decision making* 17, 2 (2023), 147–165.

[8] Qingxin Chen, Jialong Li, and Kenji Tei. 2023. Attention-guiding takeover requests for situation awareness in semi-autonomous driving. In *Companion of the 2023 ACM/IEEE International Conference on Human-Robot Interaction*. 416–421.

[9] Kashyap Chitta, Aditya Prakash, and Andreas Geiger. 2021. Neat: Neural attention fields for end-to-end autonomous driving. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 15793–15803.

[10] Andy Clark. 2013. Whatever next? Predictive brains, situated agents, and the future of cognitive science. *Behavioral and brain sciences* 36, 3 (2013), 181–204.

[11] Joseph B Claveria, Salvador Hernandez, Jason C Anderson, and Eric L Jessup. 2019. Understanding truck driver behavior with respect to cell phone use and vehicle operation. *Transportation research part F: traffic psychology and behaviour* 65 (2019), 389–401.

[12] Jacob Cohen. 2013. *Statistical power analysis for the behavioral sciences*. routledge.

[13] Jacob Cohen, Patricia Cohen, Stephen G West, and Leona S Aiken. 2013. *Applied multiple regression/correlation analysis for the behavioral sciences*. Routledge.

[14] Mike X Cohen. 2014. *Analyzing neural time series data: theory and practice*. MIT press.

[15] Michael GH Coles. 1989. Modern mind-brain reading: Psychophysiology, physiology, and cognition. *Psychophysiology* 26, 3 (1989), 251–269.

[16] Mark Colley, Benjamin Eder, Jan Ole Rixen, and Enrico Rukzio. 2021. Effects of semantic segmentation visualization on trust, situation awareness, and cognitive load in highly automated vehicles. In *Proceedings of the 2021 CHI conference on human factors in computing systems*. 1–11.

[17] Mark Colley, Max Rädler, Jonas Glimmann, and Enrico Rukzio. 2022. Effects of scene detection, scene prediction, and maneuver planning visualizations on trust, situation awareness, and cognitive load in highly automated vehicles. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 6, 2 (2022), 1–21.

[18] Mark Colley, Oliver Speidel, Jan Strohbeck, Jan Ole Rixen, Jan Henry Belz, and Enrico Rukzio. 2023. Effects of uncertain trajectory prediction visualization in highly automated vehicles on trust, situation awareness, and cognitive load. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 7, 4 (2023), 1–23.

[19] On-Road Automated Driving (ORAD) Committee. 2021. *Taxonomy and definitions for terms related to driving automation systems for on-road motor vehicles*. SAE international.

[20] David Crundall. 2016. Hazard prediction discriminates between novice and experienced drivers. *Accident Analysis & Prevention* 86 (2016), 47–58.

[21] Luca Cultrera, Federico Becattini, Lorenzo Seidenari, Pietro Pala, and Alberto Del Bimbo. 2023. Explaining autonomous driving with visual attention and end-to-end trainable region proposals. *Journal of Ambient Intelligence and Humanized Computing* (2023), 1–13.

[22] Mary Missy Cummings. 2014. Man versus machine or man+ machine? *IEEE Intelligent Systems* 29, 5 (2014), 62–69.

[23] Carlo J De Luca. 1997. The use of surface electromyography in biomechanics. *Journal of applied biomechanics* 13, 2 (1997), 135–163.

[24] Joost CF De Winter, Riender Happee, Marieke H Martens, and Neville A Stanton. 2014. Effects of adaptive cruise control and highly automated driving on workload and situation awareness: A review of the empirical evidence. *Transportation research part F: traffic psychology and behaviour* 27 (2014), 196–217.

[25] Arnaud Delorme and Scott Makeig. 2004. EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *Journal of neuroscience methods* 134, 1 (2004), 9–21.

[26] Yuqi Deng, Robert MG Reinhart, Inyong Choi, and Barbara G Shinn-Cunningham. 2019. Causal links between parietal alpha activity and spatial auditory attention. *elife* 8 (2019), e51184.

[27] Yaohan Ding, Lesong Jia, and Na Du. 2024. One size does not fit all: Designing and evaluating criticality-adaptive displays in highly automated vehicles. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems*. 1–15.

[28] Finale Doshi-Velez and Been Kim. 2017. Towards a rigorous science of interpretable machine learning. *arXiv preprint arXiv:1702.08608* (2017).

[29] Alexey Dosovitskiy, German Ros, Felipe Codevilla, Antonio Lopez, and Vladlen Koltun. 2017. CARLA: An open urban driving simulator. In *Conference on robot learning*. PMLR, 1–16.

[30] Na Du, Feng Zhou, Elizabeth Pulver, Dawn Tilbury, Lionel P Robert, Anuj K Pradhan, and X Jessie Yang. 2020. Predicting takeover performance in conditionally automated driving. In *Extended abstracts of the 2020 chi conference on human factors in computing systems*. 1–8.

[31] Na Du, Feng Zhou, Dawn Tilbury, Lionel Peter Robert, and X Jessie Yang. 2021. Designing alert systems in takeover transitions: The effects of display information and modality. In *13th international conference on automotive user interfaces and interactive vehicular applications*. 173–180.

[32] Mica R Endsley. 1995. Measurement of situation awareness in dynamic systems. *Human factors* 37, 1 (1995), 65–84.

[33] Mica R Endsley. 2017. Toward a theory of situation awareness in dynamic systems. In *Situational awareness*. Routledge, 9–42.

[34] Alexander Feierle, Fabian Schlichtherle, and Klaus Bengler. 2021. Augmented reality head-up display: A visual support during malfunctions in partially automated driving? *IEEE Transactions on Intelligent Transportation Systems* 23, 5 (2021), 4853–4865.

[35] Xiaofeng Gao, Xingwei Wu, Samson Ho, Teruhisa Misu, and Kumar Akash. 2022. Effects of augmented-reality-based assisting interfaces on drivers' object-wise situational awareness in highly autonomous vehicles. In *2022 IEEE Intelligent Vehicles Symposium (IV)*. IEEE, 563–572.

[36] Rafael Cirino Gonçalves, Tyron Louw, Gustav Markkula, and Natasha Merat. 2019. Applicability of risky decision-making theory to understand drivers' behaviour during transitions of control in vehicle automation. In *17º Congresso Internacional de Ergonomia e Usabilidade de Interfaces Humano-Tecnologia eo 17º Congresso Internacional de Ergonomia e Usabilidade de Interfaces e Interação Humano-Computador*. Blucher, 140–154.

[37] Thomas Alexander Goodge, Frank Pollick, and Stephen Anthony Brewster. 2024. Can You Hazard a Guess?: Evaluating the Effect of Augmented Reality Cues on Driver Hazard Prediction. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems*. 1–28.

[38] Alexandre Gramfort, Martin Luessi, Eric Larson, Denis A. Engemann, Daniel Strohmeier, Christian Brodbeck, Roman Goj, Mainak Jas, Teon Brooks, Lauri Parkkonen, and Matti S. Hämäläinen. 2013. MEG and EEG Data Analysis with MNE-Python. *Frontiers in Neuroscience* 7, 267 (2013), 1–13. doi:10.3389/fnins.2013.00267

[39] Andreas Gregoriades and Maria Pampaka. 2016. Enhancing drivers' situation awareness. In *Advances in Human Aspects of Transportation: Proceedings of the AHFE 2016 International Conference on Human Factors in Transportation, July 27-31, 2016, Walt Disney World®, Florida, USA*. Springer, 301–312.

[40] Peter A Hancock. 2020. Imposing limits on autonomous systems. In *New paradigms in ergonomics*. Routledge, 134–141.

[41] Peter A Hancock, Illah Nourbakhsh, and Jack Stewart. 2019. On the future of transportation in an era of automated and autonomous vehicles. *Proceedings of the National Academy of Sciences* 116, 16 (2019), 7684–7691.

[42] Paul W Hodges and Bang H Bui. 1996. A comparison of computer-based methods for the determination of onset of muscle contraction using electromyography. *Electroencephalography and Clinical Neurophysiology/Electromyography and Motor Control* 101, 6 (1996), 511–519.

[43] Chunxi Huang, Jiyao Wang, Ange Wang, Qihao Huang, and Dengbo He. 2025. The Effect of Advanced Driver Assistance Systems on Truck Drivers' Defensive Driving Behaviors: Insights from a Preliminary On-Road Study. *International Journal of Human–Computer Interaction* (2025), 1–15.

[44] Chunxi Huang, Jiyao Wang, Song Yan, and Dengbo He. 2024. Exploring factors related to drivers' mental model of and trust in advanced driver assistance systems using an ABN-based mixed approach. *IEEE Transactions on Human-Machine Systems* (2024).

[45] Wei-Chi Huang, Lin-Han Fan, Zi-Jian Han, and Ya-Feng Niu. 2024. Enhancing safety in conditionally automated driving: Can more takeover request visual information make a difference in hazard scenarios with varied hazard visibility? *Accident Analysis & Prevention* 205 (2024), 107687.

[46] Highway Loss Data Institute. 2023. Predicted availability of safety features on registered vehicles—a 2023 update. *HLDI Bull* 40, 2 (2023).

[47] Kyung Hun Jung, Jack T Labriola, and Hyunjin Baek. 2023. Projecting the planned trajectory of a Level–2 automated vehicle in the windshield: Effects on human drivers' take–over response to silent failures. *Applied Ergonomics* 111 (2023), 104047.

[48] Dina Kanaan and Birsen Donmez. 2024. How are automation failures characterized in the driving domain? Insights from a scoping review. *Journal of Cognitive Engineering and Decision Making* 18, 4 (2024), 293–301.

[49] Alex Kendall and Yarin Gal. 2017. What uncertainties do we need in bayesian deep learning for computer vision? *Advances in neural information processing systems* 30 (2017).

[50] Gwangbin Kim, Dohyeon Yeo, Taewoo Jo, Daniela Rus, and SeungJun Kim. 2023. What and when to explain? on-road evaluation of explanations in highly automated vehicles. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 7, 3 (2023), 1–26.

[51] Moritz Körber, Christian Gold, David Lechner, and Klaus Bengler. 2016. The influence of age on the take-over of vehicle control in highly automated driving. *Transportation research part F: traffic psychology and behaviour* 39 (2016), 19–32.

[52] CY KRAMERß. 1956. Extension of multiple range tests to group means with unequal numbers of replication. *Biometrics* 12 (1956), 307–310.

[53] Okkeun Lee, Rebecca Currano, Dave Miller, Hyochang Kim, and David Sirkin. 2024. Designing Visual Signals to Support Situation Awareness Recovery in Conditional Automated Driving. In *Proceedings of the 16th International Conference on Automotive User Interfaces and Interactive Vehicular Applications*. 248–258.

[54] Sangwon Lee, Jeonguk Hong, Gyewon Jeon, Jeongmin Jo, Sanghyeok Boo, Hwiseong Kim, Seoyoon Jung, Jieun Park, Inheon Choi, and Sangyeon Kim. 2023. Investigating effects of multimodal explanations using multiple In-vehicle displays for takeover request in conditionally automated driving. *Transportation research part F: traffic psychology and behaviour* 96 (2023), 1–22.

[55] Qingkun Li, Zhenyuan Wang, Wenjun Wang, Chao Zeng, Guofa Li, Quan Yuan, and Bo Cheng. 2021. An adaptive time budget adjustment strategy based on a take-over performance model for passive fatigue. *IEEE Transactions on Human-Machine Systems* 52, 5 (2021), 1025–1035.

[56] Tyron Louw, Jonny Kuo, Richard Romano, Vishnu Radhakrishnan, Michael G Lenné, and Natasha Merat. 2019. Engaging in NDRTs affects drivers' responses and glance patterns after silent automation failures. *Transportation research part F: traffic psychology and behaviour* 62 (2019), 870–882.

[57] Scott M Lundberg and Su-In Lee. 2017. A unified approach to interpreting model predictions. *Advances in neural information processing systems* 30 (2017).

[58] Karthik Mahadevan, Sowmya Somanath, and Ehud Sharlin. 2018. Communicating awareness and intent in autonomous vehicle-pedestrian interaction. In *Proceedings of the 2018 CHI conference on human factors in computing systems*. 1–12.

[59] Scott Makeig. 1993. Auditory event-related dynamics of the EEG spectrum and effects of exposure to tones. *Electroencephalography and clinical neurophysiology* 86, 4 (1993), 283–293.

[60] Vladimir Maksimenko, Xinwei Li, Eui-Jin Kim, and Prateek Bansal. 2025. Video-Based experiments better unveil societal biases towards ethical decisions of autonomous vehicles. *Transportation Research Part C: Emerging Technologies* 179 (2025), 105284.

[61] J-B Manchon, 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 (2022), 281–295.

[62] Gustav Markkula, Richard Romano, Ruth Madigan, Charles W Fox, Oscar T Giles, and Natasha Merat. 2018. Models of human decision-making as tools for estimating and optimizing impacts of vehicle automation. *Transportation research record* 2672, 37 (2018), 153–163.

[63] Daniel McFadden. 1972. Conditional logit analysis of qualitative choice behavior. (1972).

[64] Ranjana K Mehta and Raja Parasuraman. 2013. Neuroergonomics: a review of applications to physical and cognitive work. *Frontiers in human neuroscience* 7 (2013), 889.

[65] Rhiannon Michelmore, Matthew Wicker, Luca Laurenti, Luca Cardelli, Yarin Gal, and Marta Kwiatkowska. 2020. Uncertainty quantification with statistical guarantees in end-to-end autonomous driving control. In *2020 IEEE international conference on robotics and automation (ICRA)*. IEEE, 7344–7350.

[66] John A Michon. 1985. A critical view of driver behavior models: what do we know, what should we do? In *Human behavior and traffic safety*. Springer, 485–524.

[67] Tim Miller. 2019. Explanation in artificial intelligence: Insights from the social sciences. *Artificial intelligence* 267 (2019), 1–38.

[68] Brian Ka-Jun Mok, Mishel Johns, Key Jung Lee, Hillary Page Ive, David Miller, and Wendy Ju. 2015. Timing of unstructured transitions of control in automated driving. In *2015 IEEE intelligent vehicles symposium (IV)*. IEEE, 1167–1172.

[69] Callum Mole, Jami Pekkanen, William Sheppard, Tyron Louw, Richard Romano, Natasha Merat, Gustav Markkula, and Richard Wilkie. 2020. Predicting takeover response to silent automated vehicle failures. *Plos one* 15, 11 (2020), e0242825.

[70] Alfonso Montella, Filomena Mauriello, Mariano Pernetti, and Maria Rella Riccardi. 2021. Rule discovery to identify patterns contributing to overrepresentation and severity of run-off-the-road crashes. *Accident Analysis & Prevention* 155 (2021), 106119.

[71] George Nasser, Ben W Morrison, Mark W Wiggins, and Angela Hoang. 2025. The mismatch between perceived situation awareness and hazard recognition in automated driving. *Applied Ergonomics* 128 (2025), 104562.

[72] Sina Nordhoff, Tyron Louw, Ruth Madigan, Yee Mun Lee, Satu Innamaa, Esko Lehtonen, Fanny Malin, Afsaneh Bjorvatn, Anja Beuster, Riender Happee, et al. 2022. Profiling the enthusiastic, neutral, and sceptical users of conditionally automated cars in 17 Countries: A questionnaire study. *Journal of Advanced Transportation* 2022, 1 (2022), 8053228.

[73] US NTSB. 2017. Collision between a car operating with automated vehicle control systems and a tractor-semitrailer truck near Williston, Florida, May 7, 2016.

[74] Daniel Omeiza, Helena Webb, Marina Jirotka, and Lars Kunze. 2021. Explanations in autonomous driving: A survey. *IEEE Transactions on Intelligent Transportation Systems* 23, 8 (2021), 10142–10162.

[75] Sayaka Ono, Haruto Sasaki, Hitoshi Kumon, Yoshitaka Fuwamoto, Shunsuke Kondo, Takuji Narumi, Tomohiro Tanikawa, and Michitaka Hirose. 2019. Improvement of driver active interventions during automated driving by displaying trajectory pointers—A driving simulator study. *Traffic injury prevention* 20, sup1 (2019), S152–S156.

[76] François Perrin, Jacques Pernier, Olivier Bertrand, and Jean Francois Echallier. 1989. Spherical splines for scalp potential and current density mapping. *Electroencephalography and clinical neurophysiology* 72, 2 (1989), 184–187.

[77] Gert Pfurtscheller and FH Lopes Da Silva. 1999. Event-related EEG/MEG synchronization and desynchronization: basic principles. *Clinical neurophysiology* 110, 11 (1999), 1842–1857.

[78] Angkoon Phinyomark, Rami N. Khushaba, and Erik Scheme. 2018. Feature extraction and selection for myoelectric control based on wearable EMG sensors. *Sensors* 18, 5 (2018), 1615.

[79] Marco Tulio Ribeiro, Sameer Singh, and Carlos Guestrin. 2016. " Why should i trust you?" Explaining the predictions of any classifier. In *Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining*. 1135–1144.

[80] Yves Rosseel. 2012. lavaan: An R package for structural equation modeling. *Journal of statistical software* 48, 1 (2012), 1–36.

[81] Afraiz Tariq Satti, Jiyoun Kim, Eunsurk Yi, Hwi-young Cho, and Sungbo Cho. 2021. Microneedle array electrode-based wearable EMG system for detection of driver drowsiness through steering wheel grip. *Sensors* 21, 15 (2021), 5091.

[82] Jinlei Shi, Chunlei Chai, Ruiyi Cai, Haoran Wei, Youcheng Zhou, Hao Fan, Wei Zhang, and Natasha Merat. 2024. Effects of various in-vehicle human–machine interfaces on drivers' takeover performance and gaze pattern in conditionally automated vehicles. *International Journal of Human-Computer Studies* 192 (2024), 103362.

[83] Rawan Srour Zreik, Thomas Goodge, Monika Harvey, and Stephen Brewster. 2024. Where's my TOR?: Evaluating the Effect of Take-Over Request Source on Older Drivers' Control Transition in Level 3 Cars. *Proceedings of the ACM on Human-Computer Interaction* 8, MHCI (2024), 1–25.

[84] Long Sun and Lingsen Hua. 2019. Effects of hazard types on drivers' risk rating and hazard response in a video-based hazard perception task. *PLoS one* 14, 3 (2019), e0214226.

[85] Yuxiang Sun, Weixun Zuo, and Ming Liu. 2020. See the future: A semantic segmentation network predicting ego-vehicle trajectory with a single monocular camera. *IEEE Robotics and Automation Letters* 5, 2 (2020), 3066–3073.

[86] Lei Tai, Peng Yun, Yuying Chen, Congcong Liu, Haoyang Ye, and Ming Liu. 2019. Visual-based autonomous driving deployment from a stochastic and uncertainty-aware perspective. In *2019 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2622–2628.

[87] Richard M Taylor. 2017. Situational awareness rating technique (SART): The development of a tool for aircrew systems design. In *Situational awareness*. Routledge, 111–128.

[88] Armin W Thomas, Felix Molter, and Ian Krajbich. 2021. Uncovering the computational mechanisms underlying many-alternative choice. *Elife* 10 (2021), e57012.

[89] Gregor Thut, Annika Nietzel, Stephan A Brandt, and Alvaro Pascual-Leone. 2006. α-Band electroencephalographic activity over occipital cortex indexes visuospatial attention bias and predicts visual target detection. *Journal of neuroscience* 26, 37 (2006), 9494–9502.

[90] Amir Tjolleng, Jaesik Yang, and Kihyo Jung. 2022. Analysis of leg muscle activities and foot angles while pressing the accelerator pedal by different foot

postures. *Applied Sciences* 12, 24 (2022), 13025.

[91] Koen van de Merwe, Steven Mallam, and Salman Nazir. 2024. Agent transparency, situation awareness, mental workload, and operator performance: A systematic literature review. *Human Factors* 66, 1 (2024), 180–208.

[92] Ange Wang, Jiyao Wang, Chunxi Huang, Dengbo He, and Hai Yang. 2025. Exploring how physio-psychological states affect drivers' takeover performance in conditional automated vehicles. *Accident Analysis & Prevention* 216 (2025), 108022.

[93] Ange Wang, Haohan Yang, Jiyao Wang, Hai Yang, and Dengbo He. 2026. Driver cognitive load estimation in conditional driving with aligned attention-enabled multimodal fusion. *Transportation Research Part C: Emerging Technologies* 183 (2026), 105471.

[94] Jiyao Wang, Suzan Ayas, Jiahao Zhang, Xiao Wen, Dengbo He, and Birsen Donmez. 2025. Towards generalizable drowsiness monitoring with physiological sensors: A preliminary study. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, Vol. 69. SAGE Publications Sage CA: Los Angeles, CA, 1145–1150.

[95] Jiyao Wang, Wenbo Li, Zhenyu Wang, Suzan Ayas, Birsen Donmez, Dengbo He, and Kaishun Wu. 2026. DrowsyDG-Phys: Generalizable driver drowsiness estimation in conditional automated vehicles using physiological signals. *Accident Analysis & Prevention* 228 (2026), 108407.

[96] Jiyao Wang, Hao Lu, Ange Wang, Xiao Yang, Yingcong Chen, Dengbo He, and Kaishun Wu. 2025. PhysMLE: Generalizable and Priors-Inclusive Multi-Task Remote Physiological Measurement. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 47, 6 (2025), 4908–4925. doi:10.1109/TPAMI.2025.3545598

[97] Jiyao Wang, Ran Tu, Ange Wang, and Dengbo He. 2024. Trust in range estimation system in battery electric vehicles–A mixed approach. *IEEE Transactions on Human-Machine Systems* 54, 3 (2024), 250–259.

[98] Jiyao Wang, Ange Wang, Haolong Hu, Kaishun Wu, and Dengbo He. 2024. Multi-source domain generalization for ecg-based cognitive load estimation: Adversarial invariant and plausible uncertainty learning. In *ICASSP 2024-2024 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 1631–1635.

[99] Jiyao Wang, Ange Wang, Song Yan, Dengbo He, and Kaishun Wu. 2024. Revisiting Interactions of Multiple Driver States in Heterogenous Population and Cognitive Tasks. *arXiv preprint arXiv:2412.13574* (2024).

[100] Jiyao Wang, Xiao Yang, Qingyong Hu, Jiankai Tang, Can Liu, Dengbo He, Yuntao Wang, Ying-Cong Chen, and Kaishun Wu. 2025. PhysDrive: A Multimodal Remote Physiological Measurement Dataset for In-vehicle Driver Monitoring. In *The Thirty-ninth Annual Conference on Neural Information Processing Systems*.

[101] Jiyao Wang, Xiao Yang, Zhenyu Wang, Ximeng Wei, Ange Wang, Dengbo He, and Kaishun Wu. 2024. Efficient mixture-of-expert for video-based driver state and physiological multi-task estimation in conditional autonomous driving. *arXiv preprint arXiv:2410.21086* (2024).

[102] Jiyao Wang, Wenbo Zhang, Xiao Luke Wen, Ran Tu, and Dengbo He. 2026. Modeling battery electric vehicle users' charging decisions in scenarios with both time-related and distance-related anxiety. *Travel Behaviour and Society* 42 (2026), 101111.

[103] Geoffrey F Woodman, Sisi Wang, David W Sutterer, Robert MG Reinhart, and Keisuke Fukuda. 2022. Alpha suppression indexes a spotlight of visual-spatial attention that can shine on both perceptual and memory representations. *Psychonomic Bulletin & Review* 29, 3 (2022), 681–698.

[104] Song Yan, Chunxi Huang, and Dengbo He. 2024. CH-LSTTM: A taxonomy of traffic hazards. In *2024 IEEE 4th International Conference on Human-Machine Systems (ICHMS)*. IEEE, 1–6.

A Goodness-of-fit of the SEM models

Table 6: Goodness-of-fit of the SEM models.

Goodness-of-fit measures	Fit Criteria		Model						
	Good fit	Acceptable fit	Takeover Success	Takeover Lead Time	F_θ Activation Time	P_α Activation Time	F_θ Power Ratio	P_α Power Ratio	P_α Time Lag
Chi-square/Degree of Freedom (χ^2/df)	[0,2]	(2,3)	1.559	3.882	3.721	3.714	3.837	3.722	3.727
p-value (Chi-square)	< .05	< .05	.02	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001
GFI	[0.95,1]	[0.90,0.95)	0.991	0.978	0.979	0.979	0.981	0.981	0.979
AGFI	[0.90,1]	[0.85,0.90)	0.983	0.954	0.957	0.957	0.961	0.961	0.958
CFI	[0.97,1]	[0.95,0.97)	0.979	0.973	0.889	0.888	0.884	0.888	0.888
RMSEA	[0,0.05]	(0.05,0.08)	0.056	0.067	0.123	0.123	0.126	0.123	0.123
SRMR	[0,0.05]	(0.05,0.08]	0.061	0.052	0.05	0.05	0.051	0.05	0.051

Notes: GFI: Goodness of Fit Index; AGFI: Adjusted Goodness-of-Fit Index; CFI: Comparative Fit Index; RMSEA: Root Mean Square Error of Approximation; SRMR: Standardized Root Mean Square Residual.