Safety21

INNOVATING FOR SAFETY

US DOT National
University Transportation Center for Safety

Carnegie Mellon University

Penn University of Pennsylvania

The Ohio State University

Morgan State University

CCAC Community College of Allegheny County

The University of Texas Rio Grande Valley

Project Title

Planning and Policy for Safer Roads with Autonomous Vehicles: Decision Making Behavior in Dilemma-inducing Situations - Phase II

Investigators

Fatemeh Nazari (https://orcid.org/0000-0003-1587-1848)

Final Report – August 2026

DISCLAIMER

Technical Report Documentation Page

<table>
<tr><td>1. Report No.
620</td><td>2. Government Accession No.</td><td>3. Recipient's Catalog No.</td></tr>
<tr><td colspan="2">4. Title and Subtitle
Planning and Policy for Safer Roads with Autonomous Vehicles: Decision Making Behavior in Dilemma-inducing Situations - Phase II</td><td>5. Report Date
August 30, 2026
6. Performing Organization Code</td></tr>
<tr><td colspan="2">7. Author(s)
Norma Escalante, https://orcid.org/0009-0008-6780-8836
Fatemeh Nazari, https://orcid.org/0000-0003-1587-1848
Victor Contreras, https://orcid.org/0009-0007-6289-3333
Mohamadhossein Noruzoliaee, https://orcid.org/0000-0003-3860-8911</td><td>8. Performing Organization Report No.</td></tr>
<tr><td colspan="2">9. Performing Organization Name and Address
The University of Texas Rio Grande Valley (UTRGV)
1201 W University Dr
Edinburg, Texas 78539</td><td>10. Work Unit No.
11. Contract or Grant No.
Federal Grant # 69A3552344811 / 69A3552348316</td></tr>
<tr><td colspan="2">12. Sponsoring Agency Name and Address
Safety21 University Transportation Center
Carnegie Mellon University
5000 Forbes Avenue
Pittsburgh, PA 15213</td><td>13. Type of Report and Period Covered
Final Report (July 1, 2025 – July 31, 2026)
14. Sponsoring Agency Code
USDOT</td></tr>
<tr><td colspan="3">15. Supplementary Notes
Conducted in cooperation with the U.S. Department of Transportation, Federal Highway Administration.</td></tr>
<tr><td colspan="3">16. Abstract
The safe deployment of autonomous vehicles (AVs) depends not only on advances in automated driving technology but also on understanding how humans respond when confronted with complex, safety-critical situations. This project investigated human decision-making and take-over behavior in AV dilemma scenarios, where automated systems and human drivers may face competing safety priorities involving vehicle occupants and vulnerable road users. To move beyond conventional hypothetical surveys and thought experiments, the study developed a multimodal neuro-behavioral experimental framework using an interactive driving simulator. Neurophysiological responses, subjective emotional assessments, observed take-over behavior, and scenario characteristics were collected and integrated within a hierarchical Bayesian survival model to examine the occurrence and timing of driver intervention while accounting for individual heterogeneity. The results indicate that take-over behavior is influenced by both dilemma context and neurophysiological responses. Frontal alpha activity showed the strongest association with intervention timing, while scenarios involving multiple pedestrians were associated with a greater likelihood of intervention. Subjective emotional responses provided comparatively limited additional explanatory power, and substantial participant-specific variability was observed across the neuro-behavioral responses. These findings highlight the importance of considering both contextual and individual factors when evaluating human interaction with automated driving systems. The project provides a methodological foundation for integrating behavioral and physiological information into the evaluation of shared-control strategies, driver-monitoring systems, and other human-centered AV safety technologies.</td></tr>
</table>

<table>
<tr><td>17. Key Words
Autonomous vehicle; Safety; human-AV interaction; neuro-behavioral modeling; Policy analysis; AV deployment</td><td colspan="3">18. Distribution Statement
No restrictions.</td></tr>
<tr><td>19. Security Classif. (of this report)
Unclassified</td><td>20. Security Classif. (of this page)
Unclassified</td><td>21. No. of Pages
25</td><td>22. Price</td></tr>
</table>

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

Table of Contents

1. Introduction

Autonomous vehicles (AVs) are transitioning from experimental prototypes to commercial mobility services, marking a fundamental transformation in the future of transportation. Commercial deployments also have expanded rapidly in recent years, with companies such as Waymo operating fully autonomous ride-hailing services across multiple U.S. metropolitan areas (Waymo 2020, 2022, 2023) while continuing to expand domestically and internationally (Waymo 2026). The widespread adoption of AVs is expected to improve transportation safety, traffic efficiency, accessibility, and environmental sustainability through optimized driving behavior, connected vehicle technologies, and shared autonomous mobility (Fagnant and Kockelman 2015; Andreson et al. 2016; Litman 2018).

Despite these anticipated benefits, the successful integration of AVs into real-world transportation systems depends not only on technological maturity but also on the development of regulatory and policy frameworks that govern vehicle behavior in complex and safety-critical situations. Among the most challenging of these situations are unavoidable collision scenarios, in which an AV must make decisions involving competing safety objectives, such as balancing the protection of vehicle occupants against that of vulnerable road users. Unlike conventional human-driven vehicles, these decisions are delegated to the autonomous driving system itself, raising fundamental questions regarding responsibility, public acceptance, and societal trust. Moreover, in partially AVs, an additional layer of uncertainty arises from the possibility that human occupants may intervene and assume control of the vehicle, making both machine decision-making and human take-over behavior critical considerations for the safe deployment of automated driving systems.

Understanding the process of human decision-making in these situations has therefore become a central research challenge for the safe design and governance of automated driving systems. Consequently, a growing body of the relevant research aims to characterize the behavioral, cognitive, and operational factors that influence human and societal responses to AV dilemma scenarios. Empirical studies in particular have increasingly moved beyond classical trolley-type thought experiments to investigate how humans make safety-critical decisions in realistic AV contexts. Existing studies have examined these decisions from multiple perspectives, including human factors, public preferences, and industrial practice.

From a human-centered systems perspective, collaborative driving research highlights the importance of cognitive processes, such as situational awareness, trust calibration, and shared control, in supporting effective human-AV interaction and safe take-over transitions (Xing et al.

2021). Behavioral studies further discuss that public preferences are context dependent rather than governed by rigid ethical principles. Participants simultaneously consider self-preservation, harm minimization, legal responsibility, and victim characteristics when evaluating dilemma scenarios, often favoring balanced and context-sensitive decision strategies over purely utilitarian or egoistic approaches (Bergmann et al. 2018; Radun et al. 2019; Krügel and Uhl 2022; Sui 2023). In contrast, evidence from the AV industry indicates that developers place considerably greater emphasis on crash prevention, cybersecurity, liability, and regulatory compliance than on resolving philosophical dilemma scenarios directly (Martinho et al. 2021). These studies overall suggest that human responses to AV dilemmas arise from a complex interaction of cognitive, contextual, and societal factors. However, most evidence is derived from stated-preference surveys or hypothetical questionnaires, providing limited insight into the cognitive and neurophysiological processes that govern real-time intervention decisions during safety-critical driving.

One promising approach to address the limitation discussed above is measuring neurophysiological activity during decision making using sensing technologies such as electroencephalography (EEG). Unlike conventional empirical methods, which infer cognitive processes indirectly from stated preferences or observed behavior, neurophysiological measurements provide objective, real-time information about the brain activity underlying perception, attention, emotion, and decision making. Among the available neuroimaging techniques, EEG has emerged as one of the most practical approaches due to high temporal resolution, non-invasive nature, and portability. Previous studies have shown that spectral features extracted from the EEG frequency bands (i.e., theta, alpha, and beta) capture cognitive and emotional processes related to attention, workload, executive control, and affective states such as valence, arousal, and dominance.

Advances in consumer-grade EEG technology have further increased the feasibility of incorporating neurophysiological measurements into transportation research. For example, Katsigiannis and Ramzan (2018) showed that a low-cost wireless EEG system can reliably support emotion recognition despite the substantially lower spatial resolution compared with research-grade equipment. Likewise, studies of automated driving have shown that EEG signals contain information associated with driver take-over behavior, cognitive workload, and trust under different levels of vehicle automation. A growing body of research has applied EEG to understand driver behavior and human interaction with automated driving systems. For example, Zou et al (2020) showed that EEG signals contain measurable information associated with driver take-over behavior in simulated automated driving, while Xu et al (2022) reported systematic changes in brain activity under different levels of vehicle automation and corresponding variations in driver trust. More broadly, researchers have argued that neurophysiological measurements should complement conventional stated- and revealed-preference approaches because they provide objective information about cognitive processes that cannot be directly observed through questionnaires alone. These findings verify the potential of EEG for understanding driver cognition during automated driving. Nevertheless, existing studies have primarily focused on emotion recognition, workload estimation, or trust assessment, rather than real-time intervention behavior during safety-critical situations.

Only a limited number of studies have combined neurophysiological measurements with AV decision-making. For instance, Bertheau et al (2025) reported that EEG responses varied when participants passively observed AVs making legally consequential decisions, while Ju and Kim (2024) found that the context influenced drivers' willingness to resume control in video-based emergency scenarios. Although these studies reveal that the context affects both neural responses and behavioral intentions, they remain limited to passive observation or hypothetical situations and therefore do not capture the dynamic interaction between cognition, emotion, and behavior during active driving. More importantly, existing studies have largely examined neurophysiological responses, subjective evaluations, and intervention behavior independently, with little effort to integrate these complementary sources of information within a unified analytical framework. Consequently, the mechanisms linking brain activity, emotional state, contextual characteristics, and real-time take-over behavior in differentiated driving scenarios remain poorly understood.

To address the limitations identified in the existing literature, this study develops a multimodal neuro-behavioral framework for investigating human decision-making during AV dilemma scenarios. Unlike previous studies that primarily rely on hypothetical surveys, passive observations, or isolated behavioral or physiological analyses, the proposed framework integrates consumer-grade EEG, subjective emotional responses, observed take-over behavior, and contextual characteristics collected in an interactive driving simulator. A hierarchical Bayesian survival model is developed to jointly analyze the occurrence and timing of driver intervention while accounting for participant-specific heterogeneity, and participant-independent validation is performed to evaluate the generalizability of the learned neuro-behavioral representations. By integrating neurophysiological, emotional, behavioral, and contextual information within a unified probabilistic framework, this study provides new insights into the cognitive mechanisms underlying supervisory decision-making during safety-critical AV situations and establishes a foundation for the development of more human-centered, adaptive, and trustworthy autonomous driving systems. The results verify that neurophysiological responses provide valuable complementary information for explaining driver intervention behavior beyond conventional self-reported measures, while also revealing substantial individual variability in human responses to

AV dilemma scenarios. These findings underscore the importance of integrating multimodal behavioral evidence into the design of future shared-control strategies and driver-monitoring systems for AVs.

The remainder of this paper is organized as follows. Section 2 presents the study methodology. Section 3 describes the experimental dataset and data preprocessing procedures. Section 4 presents the results of the multimodal neuro-behavioral analyses and discusses their implications for AV policy and human-centered system design. Finally, Section 5 summarizes the main findings, discusses the study limitations, and outlines directions for future research.

2. Methodology

2.1. Overarching Neuro-Behavioral Modeling Framework

Each participant, indexed by $i = 1, \dots, N$, completed two sets of driving trials under two automation conditions, $c \in \{\mathrm{C1}, \mathrm{C2}\}$. Under the passive automation condition (C1), participants merely observed the AV response in the dilemma situation without the ability to intervene the AV decision. In contrast, under the active automation condition (C2), participants had the flexibility to take over the manual control anytime during the ride. Within each condition, participants experienced four dilemma scenarios, indexed by $s = 1, \dots, S$, in which the AV faces an unavoidable collision requiring a trade-off between passenger safety and one of four pedestrian types including an adult pedestrian, a child pedestrian, a group of pedestrians, and an impaired pedestrian. Consequently, each participant completed eight driving trials (two automation conditions × four dilemma scenarios). This repeated-measured experimental design enabled the effects of automation agency to be examined while exposing each participant to identical dilemma contexts under both automation conditions.

During every trial, participants' neurophysiological responses were continuously recorded using the EEG headset. For participant i in scenario s under automation condition c, signals from five electrodes (AF3, AF4, T7, T8, and Pz; Figure 5) were synchronized with the driving simulator and recorded throughout the experiment. EEG signals were temporally aligned to the hazard-onset timestamp, band-pass filtered, corrected for physiological artifacts, and normalized relative to each participant's warm-up baseline. Frequency-domain features, including theta-, alpha-, and beta-band activity together with frontal alpha asymmetry, were subsequently extracted from the processed signals. The resulting neurophysiological representation for each trial is denoted by $\mathbf{Z}_{isc} = [Z_{isc1}, \dots, Z_{iscF}]^\top$, where F shows the number of retained EEG features. Since the subsequent survival analysis is restricted to the active automation condition, the notation is simplified throughout the remainder of this section as $\mathbf{Z}_{is} \equiv \mathbf{Z}_{is,\mathrm{C2}}$.

Unlike the passive automation condition, where participants merely observe the AV behavior, the active condition additionally produces revealed intervention behavior. In particular, we measured the elapsed time between hazard onset and vehicle take-over by participant i in scenario s, which is shown by T_{is}. The corresponding take-over event is also denoted by δ_{is}, as written in Eq. (1). When no intervention occurs before the end of the available response window, T_{is} is treated as a right-censored observation. In the present study, we model take-over time rather than a binary intervention outcome, which preserves substantially more behavioral information by distinguishing participants who intervene immediately from those who continue monitoring the automation before eventually taking control.

$$\delta_{is} = \begin{cases} 1, & \textit{if participant i performs a take-over in scenario s} \\ 0, & \textit{otherwise} \end{cases} \qquad c = C2 \tag{1}$$

Following completion of the experimental trials, participants retrospectively reported their emotional experience for each dilemma scenario using the three dimensions of the valence–arousal–dominance (VAD) framework, regardless of the automation condition. The scenario-specific subjective emotional-response vector is represented as $\mathbf{E}_{is} = [V_{is}, A_{is}, D_{is}]^{\top}$, where V_{is}, A_{is}, and D_{is} denote self-reported valence, arousal, and dominance, respectively. These subjective assessments complement the EEG-derived neurophysiological measures and facilitate comparisons between perceived emotional experience and physiological activation. Participants further completed a questionnaire measuring driving-related characteristics, represented by $\boldsymbol{P}_i = [P_{i1}, P_{i2}, \dots, P_{iK}]^{\top}$, where K denotes the number of characteristics variables. These variables remain constant across all driving trials for the same participant.

2.1.1. Hierarchical Bayesian Survival Model

Take-over behavior under the active automation condition is analyzed using an exploratory hierarchical Bayesian survival (Figure 1). Given the limited sample size, our focus is not to estimate definitive population-level effects. Rather, the objective is to investigate the exploratory link between participant characteristics, neurophysiological responses, subjective emotional assessments, dilemma context, and revealed intervention behavior. Let $S_{is}(t)$ denote the survival function, representing the probability that participant i has not yet initiated a take-over by time t, as written in Eq. (2). The corresponding hazard function (Eq. (3)), represents the instantaneous probability of initiating a take-over at time t, conditional on the participant not having intervened previously. The Kaplan-Meier estimator is first used to estimate the empirical survival functions across dilemma scenarios before fitting the survival model.

The take-over hazard for participant i in scenario s is then modeled as in Eq. (4), where $h_0(t)$ denotes the baseline hazard function. The vectors of coefficients $\boldsymbol{\beta}_P$, $\boldsymbol{\beta}_Z$, $\boldsymbol{\beta}_E$, and $\boldsymbol{\beta}_S$ are respectively associated with participant characteristics, neurophysiological responses, subjective emotional responses, and dilemma scenarios. The participant-specific random effect ($u_i \sim \mathcal{N}(0, \sigma_u^2)$) accommodates the repeated-measures structure of the experiment by allowing multiple observations from the same participant to share a common baseline intervention propensity while accounting for within-participant correlation.

For each covariate j, the hazard ratio is computed as in Eq. (5), where $HR_j > 1$ indicates an increased instantaneous take-over hazard (earlier intervention), whereas $HR_j < 1$ indicates a reduced instantaneous hazard (later intervention). Regularizing prior distributions are adopted to improve estimation stability, and posterior summaries are used to evaluate the direction, magnitude, and uncertainty of candidate associations rather than to make confirmatory population-level inferences. Consequently, credible intervals are interpreted as measures of posterior uncertainty, and coefficients whose credible intervals exclude zero are regarded as suggestive evidence warranting further investigation in larger studies.

$$S_{is}(t) = P(T_{is} > t) \tag{2}$$

$$h_{is}(t) = \lim_{\Delta t \to 0} \frac{P(t \leq T_{is} < t + \Delta t \mid T_{is} \geq t)}{\Delta t} \tag{3}$$

$$h_{is}(t) = h_0(t) \exp(\boldsymbol{\beta}_P^\top \boldsymbol{P}_i + \boldsymbol{\beta}_Z^\top \boldsymbol{Z}_{is} + \boldsymbol{\beta}_E^\top \boldsymbol{E}_{is} + \boldsymbol{\beta}_S^\top \boldsymbol{S}_s + u_i), \qquad c = C2 \tag{4}$$

$$HR_j = \exp(\beta_j) \tag{5}$$

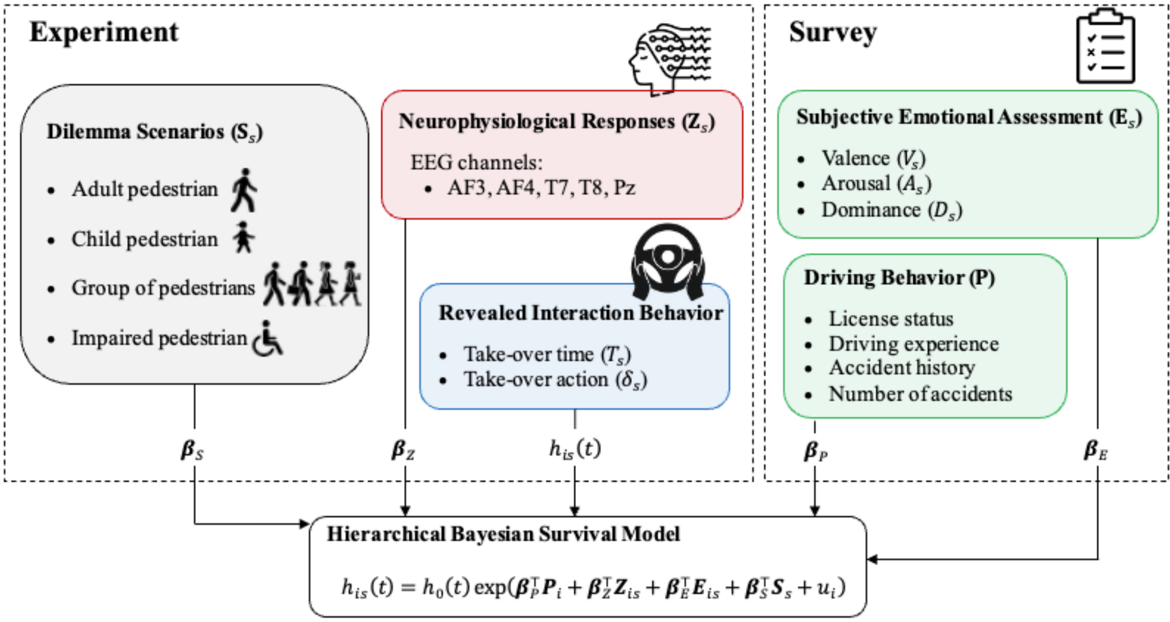


Figure 1 Proposed multimodal neuro-behavioral framework integrating experimental, neurophysiological, behavioral, and survey data within a hierarchical Bayesian survival model for analyzing driver take-over behavior during av dilemma scenarios

2.1.2. Emotional-State Prediction and Participant-Level Generalization

A complementary modeling task evaluates whether the multimodal neuro-behavioral representation contains sufficient information to predict participants' self-reported emotional responses to AV dilemma scenarios. Specifically, the scenario-level EEG features together with the corresponding dilemma scenario are used to predict the subjective emotional assessment in the VAD space. Let $\hat{\mathbf{E}}_{is} = g_\phi(\mathbf{Z}_{is}, \mathbf{S}_s)$, where $g_\phi(\cdot)$ denotes the prediction model parameterized by ϕ, and $\hat{\mathbf{E}}_{is}$ represents the predicted VAD responses for participant i in scenario s. In addition to continuous prediction, each VAD dimension is converted into predefined high- and low-state categories to enable direct comparison with conventional EEG-based emotion-recognition studies using binary classification metrics.

Since each participant contributes multiple observations, conventional random train-test splits would allow samples from the same individual to appear in both the training and testing sets, leading to overly optimistic estimates of predictive performance. To evaluate participant-independent generalization, model performance is therefore assessed primarily using leave-one-participant-out (LOPO) cross-validation. During each iteration, all observations associated with one participant are withheld as the test set, while the prediction model is trained using data from the remaining $N-1$ participants. The procedure is repeated until every participant has served as the held-out test participant exactly once. Consequently, LOPO evaluates the more challenging and practically relevant question of whether neuro-behavioral relationships learned from one group of individuals can generalize to a previously unseen participant, rather than simply predicting additional observations from participants already represented during training.

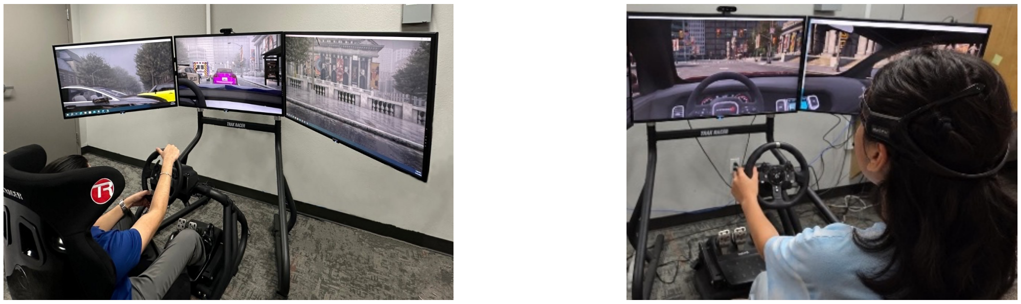
For comparison, two conventional within-participant validation strategies, which include 10-fold cross-validation and leave-one-trial-out cross-validation, are also evaluated. Unlike LOPO, these strategies permit observations from the same participant to appear in both the training and testing sets and therefore primarily assess within-participant prediction rather than participant-independent generalization. Comparing these validation schemes provides insight into the extent to which the learned neuro-behavioral representations are participant-specific versus transferable across individuals.

Overall, the proposed framework jointly models participant characteristics, neurophysiological responses, subjective emotional assessments, and revealed intervention behavior, providing complementary perspectives on human responses to AV dilemma situations. This multimodal representation enables the analysis to investigate not only whether different dilemma scenarios influence take-over behavior, but also how individual physiological and emotional responses relate to those behavioral decisions. Importantly, the framework is designed to identify associations and experimentally induced contrasts rather than to establish comprehensive causal relationships. The randomized presentation of dilemma scenarios and the controlled manipulation of intervention opportunity provide a basis for causal interpretation of the experimental contrasts between scenarios and automation conditions. In contrast, relationships involving participant characteristics, neurophysiological responses, subjective emotional assessments, and take-over behavior are interpreted as conditional associations unless additional causal identification assumptions are explicitly imposed. This distinction ensures that the conclusions remain consistent with both the experimental design and the exploratory nature of the study.

2.2. Neuro-Behavioral Experiment

2.2.1. Experimental Platform

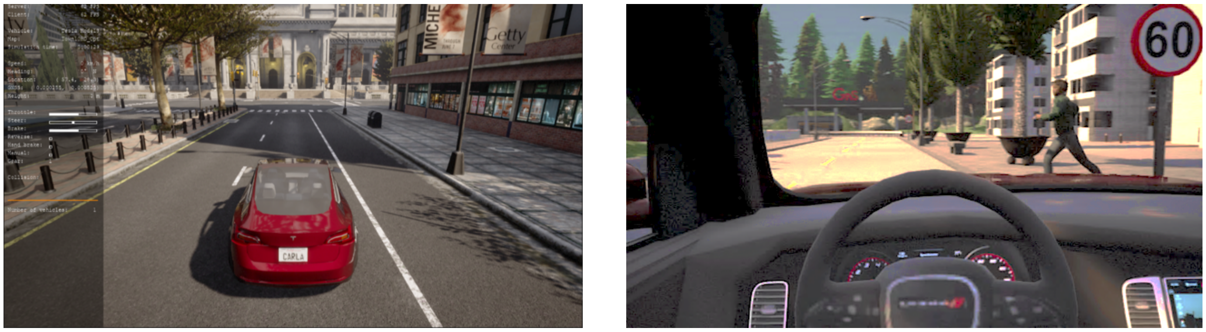
The neuro-behavioral experiment was conducted using a driving simulator consisting of a steering wheel, pedal system, and multi-monitor display that provided participants with an immersive driving environment (Figure 2(a)). The simulation environment was developed in CARLA, an open-source, high-fidelity simulator widely used for autonomous driving and intelligent transportation systems research (Dosovitskiy et al. 2017). CARLA is also increasingly adopted for human-in-the-loop automated driving research such as in Araluce et al (2022) that examine driver take-over behavior by integrating gaze information with vehicle data and take-over response times. CARLA provides configurable roadway geometries, traffic participants, vehicle dynamics, and environmental conditions, enabling safety-critical driving scenarios to be reproduced consistently across participants. Its scripting capabilities ensure identical event sequences and hazard presentations, thereby minimizing uncontrolled experimental variability.



(a) Driving simulator experiment **(b) EEG data acquisition**

Figure 2 Experimental setup for synchronized driving simulation and EEG data acquisition

During each trial, participants experienced the simulated environment from a first-person driver perspective, while a third-person perspective was available to the research team for monitoring the experiment (Figure 3). The first-person interface was maintained consistently throughout the study to ensure identical visual exposure across all participants and dilemma scenarios. Neurophysiological responses were continuously recorded using an EEG headset equipped with five electrodes (AF3, AF4, T7, T8, and Pz), as shown in Figure 2(b). Continuous EEG acquisition throughout the experimental session preserved temporal continuity across trials and enabled participant-specific baseline normalization using the warm-up driving condition. Details of electrode placement, signal preprocessing, and feature extraction are presented in the following section. The driving simulator and EEG acquisition system were temporally synchronized using a common experimental timeline. CARLA generated time-stamped records of vehicle states, scenario events, hazard onset, and participant control inputs, while EEG signals were recorded concurrently. This synchronization enabled neurophysiological activity to be aligned with specific driving events and participant interventions, providing the basis for the subsequent analysis.



(a) Third-person perspective **(b) First-person driver perspective**

Figure 3 Visualization of the CARLA-based autonomous driving simulation environment

2.2.2. Dilemma Scenario Design

Following EEG initialization, participants completed a warm-up driving session to familiarize themselves with the simulator and vehicle controls while establishing a participant-specific physiological baseline for subsequent EEG normalization. The warm-up was conducted in the same roadway environment as the experimental trials but contained no hazards or critical events. Next, participants experienced four dilemma scenarios (Figure 4), each involving an unavoidable collision with a different type of vulnerable road user: (i) an adult pedestrian, (ii) a child pedestrian, (iii) a group of pedestrians, and (iv) an impaired pedestrian. These victim categories were selected based on previous studies demonstrating that age, group size, and physical condition influence human judgments in AV dilemmas (Awad et al. 2018; Nazari and Noruzoliaee 2025). Unlike questionnaire-based studies, the present experiment embedded these dilemmas within an immersive, time-constrained driving environment, enabling simultaneous observation of participants' behavioral, emotional, and neurophysiological responses.

To isolate the effect of victim type, all other experimental conditions were held constant across scenarios, including roadway geometry, vehicle trajectory and speed, environmental conditions, hazard location, and event timing. In each trial, the AV traveled at a constant speed along a straight urban roadway. Approximately 4 s before the potential collision (time-to-collision), the designated pedestrian emerged from an initially occluded position and entered the vehicle's path. Hazard onset was defined as the instant at which the pedestrian became visible to the participant and served as the common temporal reference for aligning simulator events, EEG recordings, and behavioral responses.

Throughout each scenario, the automated driving system maintained its speed and trajectory without braking or performing evasive maneuvers. Under the active automation condition (C2), participants were instructed to monitor the roadway and assume manual control whenever they considered intervention necessary. A take-over event and the corresponding response time relative to hazard onset were recorded whenever manual control was initiated. Trials in which no intervention occurred before the predetermined collision point were treated as right-censored observations in the subsequent survival analysis.

The experimental design was intended to induce a controlled conflict between reliance on vehicle automation and the impulse to intervene in response to an imminent hazard. By varying only the pedestrian type while maintaining identical roadway geometry, vehicle dynamics, hazard visibility, and available response time, the study minimizes potential confounding from differences in driving-task difficulty. Consequently, observed differences in EEG activity, take-over behavior, and post-scenario emotional responses can be attributed primarily to the experimentally manipulated victim category.

(a) Scenario 1 – Adult pedestrian

(b) Scenario 2 – Child pedestrian

(c) Scenario 3 – Group of pedestrians

(d) Scenario 4 – Impaired pedestrian

Figure 4 AV dilemma scenarios used in the neuro-behavioral driving simulator experiment

3. Data Analysis

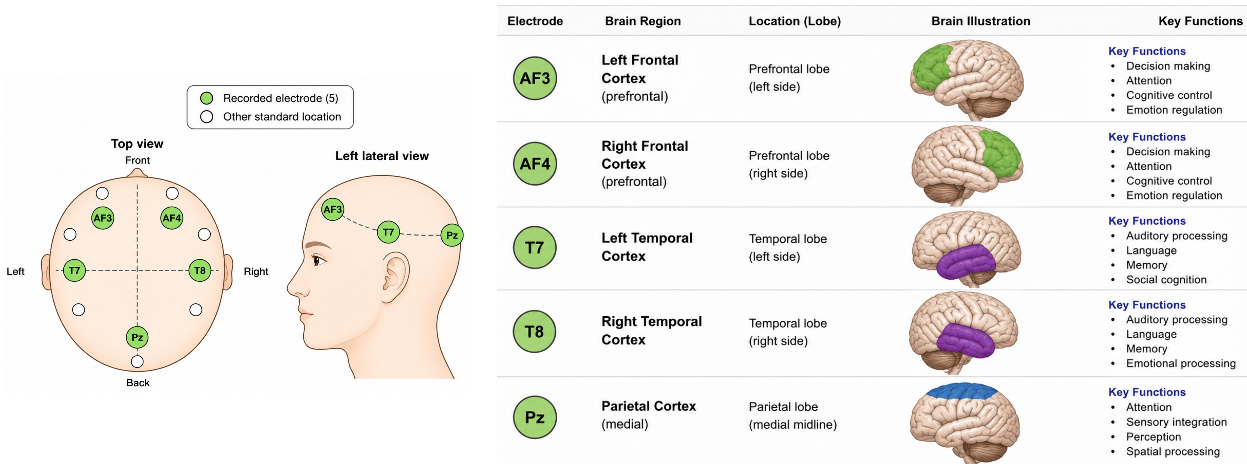
3.1. Participant Characteristics

Following the driving experiment, participants completed a questionnaire assessing their driving behavior. The questionnaire collected information on driver's license status, years of driving experience, previous crash involvement, number of crashes experienced, and whether any crash resulted in property damage or personal injury. These participant-level characteristics ($\boldsymbol{P}_i$) were used to describe the study sample and to represent driving behavior in the subsequent modeling framework.

3.2. EEG Preprocessing and Neurophysiological Feature Extraction

3.2.1. EEG Preprocessing and Feature Extraction

EEG signals were continuously recorded throughout the experiment using a five-channel headset (AF3, AF4, T7, T8, and Pz) at a sampling frequency of 128 Hz (Figure 5). Prior to analysis, the recordings were processed using a standardized preprocessing pipeline. Continuous EEG signals were segmented into scenario-specific epochs using the hazard-onset timestamp generated by the CARLA simulator as the temporal reference. Each epoch was band-pass filtered (0.5–40 Hz), power-line interference was removed using a 60 Hz notch filter, and motion- and blink-related artifacts were corrected. The resulting signals were baseline-normalized using the participant-specific warm-up session to reduce inter-individual variability. Frequency-domain features were then extracted from the theta (4–7 Hz), alpha (8–12 Hz), and beta (13–30 Hz) bands. A compact set of eight neurophysiological features, including spectral band-power measures and frontal alpha asymmetry, was retained for subsequent modeling. Frontal alpha asymmetry quantifies hemispheric asymmetry in frontal cortical activity and has been widely associated with emotional valence and approach-withdrawal behavior. Finally, all features were normalized on a participant-by-participant basis, yielding the participant- and scenario-specific neurophysiological feature vector, $\mathbf{Z}_{is}$, used in the subsequent multimodal analyses.



(a) Electrode locations (10-20 system)

(b) Underlying brain regions and key function

Figure 5 EEG electrode locations, underlying brain regions, and representative EEG signals

3.2.2. Neurophysiological Responses and Intervention Behavior

A secondary, model-independent analysis compared participants' neurophysiological responses between the passive (C1) and active (C2) automation conditions. For each participant, the core EEG features were aggregated across the four dilemma scenarios under each condition, and paired Wilcoxon signed-rank tests were used to evaluate C1–C2 differences without assuming normality. As shown in Figure 6, no statistically significant differences were observed between the two automation conditions for the examined EEG band-power measures or frontal alpha asymmetry. However, participant-level trajectories revealed substantial inter-individual variability, with increases in neural activity for some participants and decreases or minimal changes for others. This heterogeneity was particularly evident in the impaired-pedestrian scenario, where a pronounced increase in beta power for one participant resulted in greater between-subject dispersion. Overall, the results indicate that active agency does not elicit a consistent neurophysiological response across participants. Instead, the observed variability suggests that responses to intervention opportunities are highly participant-dependent, supporting the participant-specific normalization and the subsequent hierarchical multimodal modeling framework rather than relying solely on pooled group-level comparisons.

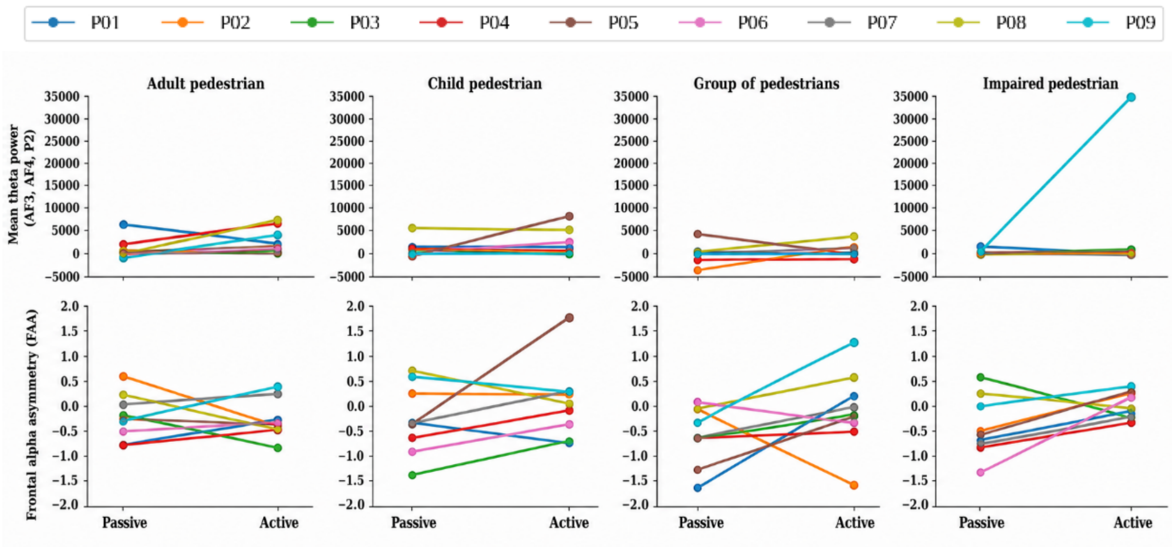


Figure 6 Participant-level changes in EEG features between passive (C1) and active (C2) automation conditions across the four AV dilemma scenarios

3.3. Subjective Emotional Responses

Participants' subjective emotional responses were assessed using the Self-Assessment Manikin, which measures valence, arousal, and dominance. Following completion of the driving experiment, participants retrospectively evaluated each dilemma scenario. Valence was rated on a scale from −5 (very negative) to +5 (very positive), arousal from −5 (very calm) to +5 (very activated), and dominance from 0 (very helpless) to +5 (fully in control). The scenario-specific VAD responses, denoted by $\mathbf{E}_{is} = [V_{is}, A_{is}, D_{is}]^{\top}$, were paired with the corresponding EEG features and behavioral observations in the subsequent multimodal analyses.

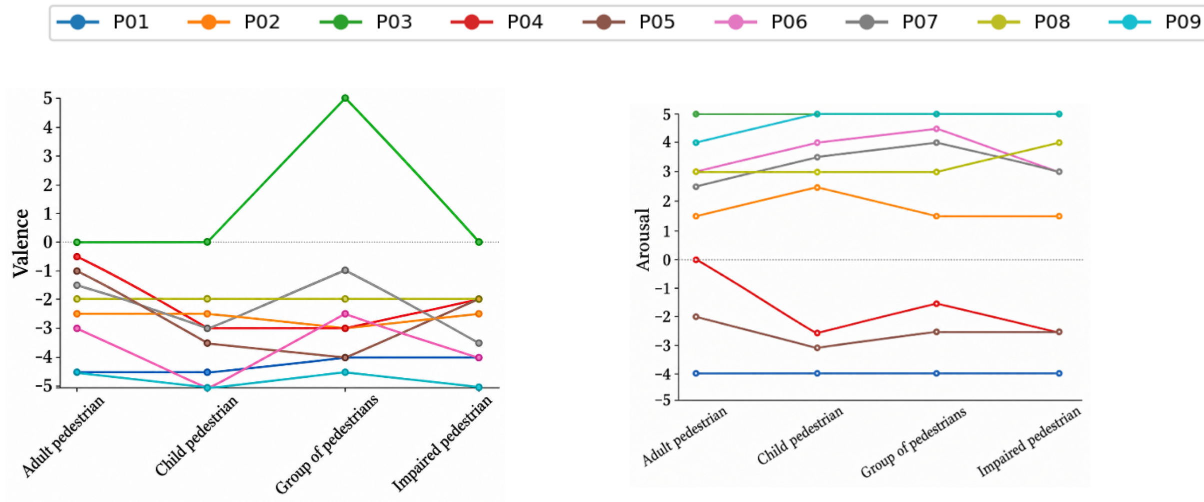
Table 1 summarizes the descriptive statistics for the four dilemma scenarios. Across all scenarios, participants reported predominantly negative valence, moderate emotional activation, and relatively low perceived control. The child-pedestrian scenario elicited the most negative emotional response (mean valence = −3.17) and the lowest perceived control (mean dominance = 0.78), whereas mean arousal varied only modestly across scenarios. Despite these overall trends, substantial variability was observed within every scenario, indicating considerable inter-individual differences in emotional appraisal. Overall, the VAD responses suggest that the AV dilemma scenarios were generally perceived as emotionally negative, moderately arousing, and associated with limited control. The pronounced participant-level variability further motivates integrating subjective emotional responses with neurophysiological and behavioral measures in the subsequent multimodal modeling framework.

Table 1. Descriptive Statistics of Post-Scenario Emotional Responses by Dilemma Scenario

Construct	Survey Question	Response Scale	Scenario	Mean ± SD	Median	Range
Valence	How positive or negative did you feel during that scenario?	−5 (Very negative) to +5 (Very positive)	Adult pedestrian	-2.17 ± 2.48	-2.50	[-5, 5]
			Child pedestrian	-3.17 ± 2.50	-4.00	[-5, 5]
			Group of pedestrians	-2.11 ± 3.07	-3.50	[-5, 5]
			Impaired pedestrian	-2.78 ± 2.53	-4.00	[-5, 5]
Arousal	How calm or activated did you feel during that scenario?	−5 (Very calm) to +5 (Very activated)	Adult pedestrian	+1.44 ± 2.89	+2.00	[-4, 5]
			Child pedestrian	+1.50 ± 3.55	+3.00	[-4, 5]
			Group of pedestrians	+1.67 ± 3.46	+2.50	[-4, 5]
			Impaired pedestrian	+1.39 ± 3.43	+2.50	[-4, 5]
Dominance	How in control or helpless did you feel during that scenario?	0 (Very helpless) to +5 (Fully in control)	Adult pedestrian	+1.06 ± 1.06	+1.00	[0, 3]
			Child pedestrian	+0.78 ± 0.88	+0.50	[0, 2]
			Group of pedestrians	+1.06 ± 1.26	+1.00	[0, 4]
			Impaired pedestrian	+0.89 ± 1.02	+0.50	[0, 3]

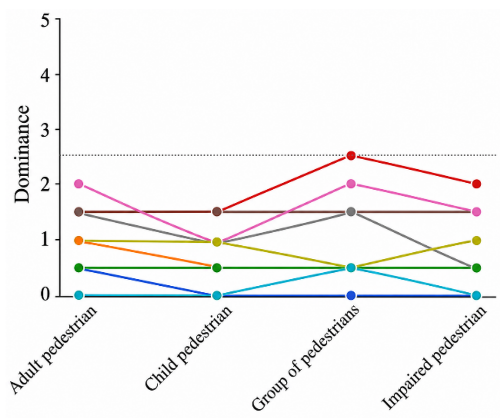
3.3.1. Participant Heterogeneity in Emotional Responses

While the descriptive statistics in Table 1 summarize the average emotional responses across scenarios, Figure 7 illustrates the corresponding participant-level trajectories for valence, arousal, and dominance. Each line represents one participant, preserving the repeated-measures structure of the experiment and highlighting within-participant changes across the four dilemma scenarios. Figure 7 reveals substantial inter-individual variability in all three emotional dimensions. Valence was generally negative across participants, although the magnitude and direction of scenario-specific changes varied considerably. Arousal exhibited a somewhat more consistent pattern, with several participants reporting increased activation in the child-pedestrian scenario relative to the adult-pedestrian scenario. However, notable floor- and ceiling-level responses were observed for a few participants. Dominance remained relatively low overall, but participants differed in how their perceived level of control changed across scenarios. Overall, the participant-level trajectories reveal that the emotional impact of the AV dilemmas cannot be fully characterized by pooled averages alone. The substantial heterogeneity observed across individuals supports the use of participant-specific and hierarchical modeling approaches that preserve within-participant variation in the subsequent multimodal analyses.



(a) Valence

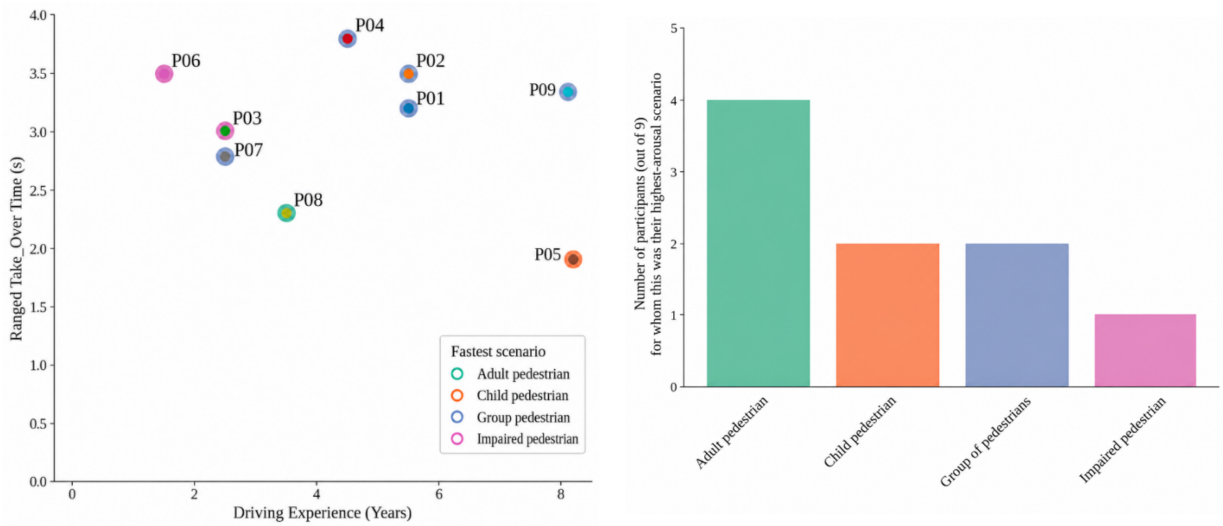
(b) Arousal



(c) Dominance

Figure 7 Participant-level trajectories of emotional responses across AV dilemma scenarios

Participant heterogeneity is further analyzed using two complementary measures with results show in Figure 8. The results show substantial differences in take-over-time variability across participants, with the scenario eliciting the fastest intervention differing from one participant to another. Similarly, no single dilemma scenario consistently produces the highest emotional arousal across participants. Although the adult-pedestrian scenario most frequently elicits the maximum arousal, all four scenarios are identified as the most emotionally activating by at least one participant. Overall, these findings reveal that sensitivity to AV dilemma scenarios varies considerably across individuals in both revealed behavior and subjective emotional responses. This participant-specific heterogeneity further supports the subsequent multimodal framework, which jointly models participant characteristics, scenario context, neurophysiological responses, emotional appraisal, and intervention behavior.



(a) Variability in take-over timing across scenarios by driving experience and fastest-response scenario

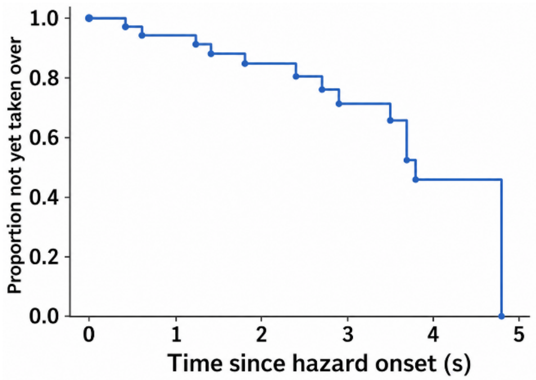
(b) Distribution of scenarios eliciting each participant's maximum self-reported arousal

Figure 8 Participant-specific sensitivity to AV dilemma scenarios

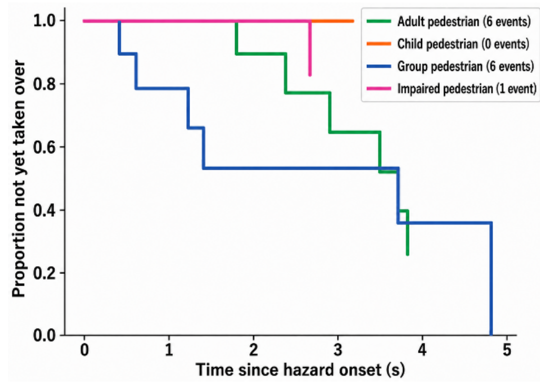
3.4. Revealed Intervention Behavior

3.4.1. Take-Over Occurrence and Timing

Take-over behavior under the active automation condition (C2) is analyzed as a time-to-event process, where the event corresponded to participant-initiated take-over and trials without intervention are treated as right-censored at the end of the response window. Across the 36 C2 trials, 13 take-over events and 23 censored observations were recorded, corresponding to an overall intervention rate of 36%. Figure 9(a) presents the Kaplan–Meier estimate of the probability of remaining under automated control following hazard onset. The survival probability decreases progressively throughout the response window, indicating that take-over decisions occurred over time rather than immediately after hazard appearance. This temporal pattern highlights the importance of modeling intervention as a time-to-event process rather than a binary outcome. Scenario-specific Kaplan–Meier curves (Figure 9(b)) suggest differences in intervention behavior across victim types. The group-of-pedestrians scenario exhibit the earliest and most frequent interventions, whereas only one take-over occurs in the impaired-pedestrian scenario and none in the child-pedestrian scenario. Given the limited number of participants and events, these scenario-specific differences should be interpreted as exploratory.



(a) **Overall survival function**

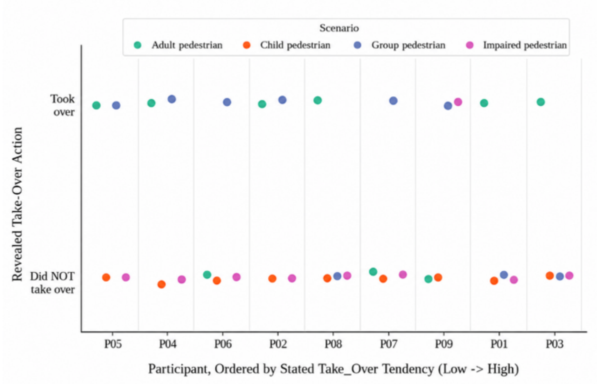


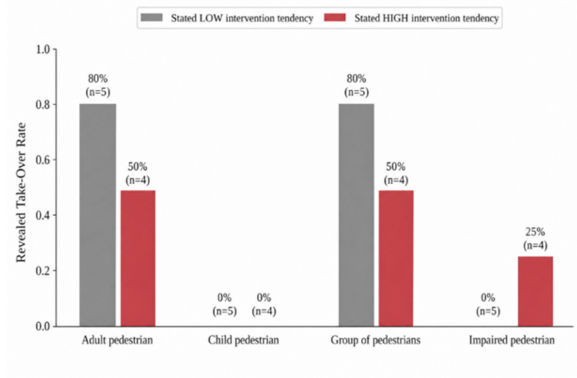
(b) Scenario-specific survival functions

Figure 9 Kaplan-Meier estimates of time to participant take-over under the active autonomation condition (C2)

3.4.2. Stated versus Revealed Take-Over Behavior

Participants' revealed take-over behavior during the active automation condition (C2) is further compared with their stated intervention tendency, which was reported in after-experiment survey. This comparison, which is shown in Figure 10, examines whether participants who reported a greater willingness to override an AV were more likely to intervene when confronted with an imminent collision. As shown in Figure 10(a), stated intervention tendency cannot consistently predict observed take-over behavior. Participants with both low and high stated intervention scores exhibited mixtures of take-over and non-take-over decisions across the four dilemma scenarios. Similarly, the scenario-specific intervention rates in Figure 10(b) reveal no clear monotonic relationship between stated preference and revealed behavior. Instead, intervention behavior varied across both participants and dilemma contexts. These findings suggest that pre-experiment attitudes alone provide an incomplete representation of decision-making in safety-critical AV situations. Rather, revealed intervention behavior appears to reflect the combined influence of participant characteristics, scenario context, and neurophysiological responses, motivating the subsequent multimodal modeling framework.





(a) Stated take-over action

(b) Scenario-specific take-over rates stratified by stated intervention tendency

Figure 10 Stated intervention tendency versus revealed take-over behavior under the active automation condition (C2)

4. Multimodal Neuro-Behavioral Modeling

4.1. Exploratory Hierarchical Bayesian Survival Analysis of Take-Over Timing

The descriptive Kaplan–Meier analysis is extended using an exploratory hierarchical Bayesian survival model to jointly analyze the occurrence and timing of participant take-over while accounting for repeated observations from the same individuals. The model was estimated using the 36 active-condition (C2) trials, comprising 13 take-over events and 23 right-censored observations. Driving behavior, neurophysiological responses, subjective emotional assessments, and dilemma scenarios were included as fixed effects, with participant-specific random effects accounting for repeated measures. Posterior sampling exhibit satisfactory convergence ($\hat{R} = 1.00$ for all parameters, effective sample sizes >3,700, and no divergent transitions), indicating efficient exploration of the posterior distribution and stable estimation of posterior summaries. The posterior estimates are summarized in Table 2.

Among **the neurophysiological variables**, frontal alpha activity exhibits the strongest association with intervention timing. Higher $AF3_{\alpha}$ activity is associated with an increased take-over hazard (HR = 2.48), whereas higher $AF4_{\alpha}$ activity is associated with a reduced hazard (HR = 0.35). The 89% credible intervals for both coefficients exclude zero. In contrast, beta-band activity shows substantial posterior uncertainty. The opposite effects observed for AF3 and AF4 suggest a lateralized relationship between prefrontal neural activity and supervisory decision-making during AV dilemma scenarios. As the prefrontal cortex is critically involved in executive control, attention, and action selection, these findings indicate that left and right prefrontal processes may contribute differently to the timing of human intervention in safety-critical situations. **Scenario context** also influences intervention behavior. Relative to the adult-pedestrian scenario, the group-of-pedestrians scenario is found to be associated with a substantially higher take-over hazard (HR = 2.96), consistent with the earlier Kaplan–Meier analysis. The child- and impaired-pedestrian scenarios exhibit lower estimated hazards, although their credible intervals included zero. **Driving experience** and **the subjective emotional responses** (valence, arousal, and dominance) likewise show no clear associations with intervention timing after accounting for the remaining model covariates.

Table 2. Posterior Estimates of Fixed Effects from the Hierarchical Bayesian Survival Model for Take-Over Timing

Covariate	Posterior Mean	SD	Hazard Ratio	89% Credible Interval	Credible Interval Excludes Zero
Driving Behavior (β_P)					
Driving experience (standardized)	-0.160	0.440	0.852	[-0.850, 0.550]	No
Neurophysiological Responses (β_Z)					
$AF3_\alpha$	**0.908**	**0.516**	**2.479**	**[0.088, 1.734]**	**Yes**
$AF3_\beta$	0.351	0.588	1.420	[-0.600, 1.270]	No
$AF4_\alpha$	**-1.055**	**0.537**	**0.348**	**[-1.907, -0.186]**	**Yes**
$AF4_\beta$	0.560	0.515	1.751	[-0.266, 1.376]	No
Subjective Emotional Responses (β_E)					
Valence	0.116	0.135	1.123	[-0.095, 0.331]	No
Arousal	-0.069	0.123	0.935	[-0.264, 0.127]	No
Dominance	0.299	0.335	1.349	[-0.230, 0.825]	No
Dilemma Scenario (β_S)					
Child pedestrian	-0.913	0.804	0.401	[-2.253, 0.328]	No
Impaired pedestrian	-1.021	0.728	0.360	[-2.250, 0.119]	No
Group of pedestrians	**1.085**	**0.610**	**2.959**	**[0.105, 2.057]**	**Yes**

Note: The reference dilemma scenario is adult pedestrian.

4.2. Participant-Independent Generalization of Neuro-Behavioral Representations

To evaluate whether the learned neuro-behavioral representations generalized beyond the training participants, the continuous VAD responses are converted into binary classes using predefined thresholds. Because the resulting class distributions are imbalanced, model performance is evaluated using balanced accuracy and F1-score in addition to conventional accuracy. Participant-independent generalization is assessed using LOPO cross-validation, while two trial-level validation schemes (10-fold cross-validation and leave-one-trial-out) are included for comparison.

The classification results are presented in Table 3. Under LOPO validation, the learned representations exhibit limited generalization across all three emotional dimensions. Balanced accuracies are near chance level for valence (49.6%), arousal (26.0%), and dominance (48.5%), indicating that the relationships learned from one group of participants transferred poorly to previously unseen individuals. The dominance classifier further shows the importance of evaluating imbalanced datasets using balanced accuracy rather than conventional accuracy, as its high overall accuracy (88.9%) primarily reflects the majority-class distribution rather than meaningful discrimination. In contrast, trial-level validation produces substantially higher performance, particularly for arousal, where balanced accuracy increases to 98.0% under 10-fold cross-validation and 95.7% under leave-one-trial-out validation. Since these validation schemes allow observations from the same participant to appear in both the training and testing sets, they primarily assess within-participant prediction rather than participant-independent generalization.

Overall, the results indicate that the learned neuro-behavioral representations primarily capture participant-specific patterns. While they accurately distinguish emotional states for previously observed participants, they do not yet generalize reliably to unseen individuals. These findings highlight the importance of participant-independent evaluation for EEG-based affective modeling and suggest that larger, more diverse datasets or participant-specific adaptation strategies will be necessary for robust real-world deployment.

Table 3. Binary Classification Performance of Neuro-Behavioral Representations for Predicting Valence, Arousal, and Dominance (VAD)

Validation Scheme	VAD Dimension	Accuracy (%)	Balanced Accuracy (%)	Majority Baseline (%)	F1-Score
Participant-Independent Validation					
LOPO	Valence	76.40	49.60	81.90	0.11
	Arousal	36.10	26.00	69.40	0.53
	Dominance	88.90	48.50	91.70	0.00
Trial-Level Validation					
10-fold CV	Valence	84.70	63.70	81.90	0.42
	Arousal	97.20	98.00	69.40	0.98
	Dominance	90.30	49.20	91.70	0.00
Leave-one-trial-out	Valence	86.10	67.50	81.90	0.50
	Arousal	95.80	95.70	69.40	0.97
	Dominance	91.70	50.00	91.70	0.00

Note: LOPO = leave-one-participant-out cross-validation.

4.3. Policy Implications

The findings provide several implications for the design, evaluation, and regulation of human-supervised AVs. Although the present study is exploratory, the results reveal that intervention behavior in safety-critical situations is influenced by a combination of neurophysiological state, dilemma context, and participant-specific characteristics rather than by a single behavioral or demographic factor. **First**, the strong participant-specific variability observed in both the hierarchical survival model and the participant-independent validation suggests that human supervisory behavior cannot be adequately represented by a single population-level model. Existing AV testing and certification procedures generally assume that drivers exhibit similar responses to takeover requests. In contrast, the present findings indicate that intervention timing and emotional responses differ substantially across individuals. Consequently, future shared-control systems should incorporate adaptive or personalized strategies that account for individual differences in cognitive state, decision-making, and intervention behavior.

Second, the significant influence of dilemma context on take-over timing highlights the importance of context-aware human-AV interaction. Participants were considerably more likely to intervene in scenarios involving multiple pedestrians than in the reference scenario, suggesting that users respond not only to collision risk but also to the perceived consequences of the situation. Rather than relying on fixed warning thresholds or uniform intervention logic, future AVs could dynamically adjust driver alerts, information presentation, or authority-sharing according to the severity and contextual characteristics of the developing hazard.

Third, the association between frontal alpha activity and intervention timing demonstrates the potential value of physiological sensing in next-generation driver monitoring systems. Current monitoring technologies primarily assess visual attention or physical engagement. Incorporating neurophysiological indicators could provide additional information regarding cognitive readiness for intervention, enabling more responsive transitions between automated and manual control. Although EEG is not yet practical for widespread deployment, these findings motivate continued research into physiological indicators that can reliably infer driver readiness.

Fourth, the poor participant-independent generalization emphasizes the need for more rigorous evaluation standards for affect-aware intelligent transportation systems. The large performance gap between leave-one-participant-out and trial-level validation demonstrates that models trained on previously observed users may perform well during development but fail to generalize to new drivers. Consequently, future regulatory and benchmarking frameworks should require participant-independent validation when evaluating driver-state monitoring, emotion-recognition, or adaptive shared-control systems to ensure that reported performance reflects real-world deployment conditions.

Finally, the results suggest that future AV policies should emphasize personalization rather than one-size-fits-all automation. Reliable deployment of neuro-behavioral decision-support systems will likely require substantially larger and more diverse training datasets together with adaptive calibration strategies that continuously learn individual behavioral patterns while maintaining safety and transparency. Such approaches may improve trust, reduce inappropriate interventions, and enable safer collaboration between human drivers and automated driving systems.

5. Conclusions

This study developed a multimodal neuro-behavioral framework to investigate human decision-making during autonomous vehicle (AV) dilemma scenarios by integrating neurophysiological responses, subjective emotional assessments, observed take-over behavior, and contextual characteristics within an interactive driving simulator. A hierarchical Bayesian survival model was used to analyze the occurrence and timing of driver intervention under safety-critical conditions.

The results indicate that take-over behavior is jointly influenced by neurophysiological activity and dilemma context. Frontal alpha activity exhibits the strongest association with intervention timing, while scenarios involving multiple pedestrians are associated with a greater propensity for intervention. In contrast, subjective emotional responses and driving experience provided limited additional explanatory power. The participant-independent validation further reveals that neuro-behavioral representations are highly individualized, highlighting the importance of adaptive approaches for future EEG-informed driver-monitoring systems.

Overall, the proposed framework demonstrates the value of integrating neurophysiological, emotional, behavioral, and contextual information to better understand human-AV interaction and to support the design of more human-centered shared-control strategies. Since this study is based on a small sample collected using a consumer-grade EEG system in a simulated driving environment, the findings need to be interpreted as exploratory. Future research should validate the framework using larger and more diverse participant populations, richer physiological sensing modalities, and more realistic driving environments.

References

1. Andreson, James M., Nidhi Karla, Karlyn D. Stanley, Paul Sorensen, Constantine Samaras, and Oluwatobi A. Oluwatola. 2016. Autonomous Vehicle Technology: A Guide for Policymakers. Retrieved from: https://www.rand.org/content/dam/rand/pubs/research_reports/rr400/rr443-2/rand_rr443-2.pdf. rr-443-2-rc. RAND Corporation.

2. Araluce, Javier, Luis M. Bergasa, Manuel Ocaña, Elena López-Guillén, Rodrigo Gutiérrez-Moreno, and J. Felipe Arango. 2022. "Driver Take-over Behaviour Study Based on Gaze Focalization and Vehicle Data in CARLA Simulator." Sensors 22 (24): 9993. https://doi.org/10.3390/s22249993

3. Awad, Edmond, Sohan Dsouza, Richard Kim, et al. 2018. "The Moral Machine Experiment." Nature 563 (7729): 59–64. https://doi.org/10.1038/s41586-018-0637-6

4. Bergmann, Lasse T., Larissa Schlicht, Carmen Meixner, et al. 2018. "Autonomous Vehicles Require Socio-Political Acceptance—An Empirical and Philosophical Perspective on the Problem of Moral Decision Making." Frontiers in Behavioral Neuroscience 12: 31. https://doi.org/10.3389/fnbeh.2018.00031

5. Bertheau, Maren A. K., Cindy Boetzel, and Christoph S. Herrmann. 2025. "Event-Related Potentials Reveal Incongruent Behavior of Autonomous Vehicles in the Moral Machine Dilemma." Scientific Reports 15 (1): 16048. https://doi.org/10.1038/s41598-025-99531-6

6. Dosovitskiy, Alexey, German Ros, Felipe Codevilla, Antonio Lopez, and Vladlen Koltun. 2017. "CARLA: An Open Urban Driving Simulator." 78: 1–16.

7. Fagnant, Daniel J., and Kara Kockelman. 2015. "Preparing a Nation for Autonomous Vehicles: Opportunities, Barriers and Policy Recommendations." Transportation Research Part A: Policy and Practice 77 (Journal Article): 167–81. https://doi.org/10.1016/j.tra.2015.04.003

8. Ju, Uijong, and Sanghyeon Kim. 2024. "Willingness to Take Responsibility: Self-Sacrifice versus Sacrificing Others in Takeover Decisions during Autonomous Driving." Heliyon 10 (9). https://doi.org/10.1016/j.heliyon.2024.e29616

9. Katsigiannis, Stamos, and Naeem Ramzan. 2018. "DREAMER: A Database for Emotion Recognition through EEG and ECG Signals from Wireless Low-Cost off-the-Shelf Devices." IEEE Journal of Biomedical and Health Informatics 22 (1): 98–107. https://doi.org/10.1109/JBHI.2017.2688239

10. Krügel, Sebastian, and Matthias Uhl. 2022. "Autonomous Vehicles and Moral Judgments under Risk." Transportation Research Part A: Policy and Practice 155: 1–10. https://doi.org/10.1016/j.tra.2021.10.016

11. Litman, Todd Alexander. 2018. Autonomous Vehicle Implementation Predictions: Implications for Transport Planning. Retrieved from: https://www.vtpi.org/avip.pdf. Victoria Transport Policy Institute.

12. Martinho, Andreia, Nils Herber, Maarten Kroesen, and Caspar Chorus. 2021. "Ethical Issues in Focus by the Autonomous Vehicles Industry." Transport Reviews, 1–22. https://doi.org/10.1080/01441647.2020.1862355

13. Nazari, Fatemeh, and Mohamadhossein Noruzoliaee. 2025. Planning and Policy for Safer Roads with Autonomous Vehicles: Moral Decision Making Behavior in Dilemma-Inducing Situations. Carnegie Mellon University. Traffic21 Institute. Safety21 University Transportation Center (UTC), U.S. Department of Transportation. https://rosap.ntl.bts.gov/view/dot/86173

14. Radun, Igor, Jenni Radun, Jyrki Kaistinen, Jake Olivier, Göran Kecklund, and Töres Theorell. 2019. "Endangering Yourself to Save Another: A Real Life Ethical Dilemma." Transportation Research Part F: Traffic Psychology and Behaviour 64: 318–22. https://doi.org/10.1016/j.trf.2019.05.015

15. Sui, Tingting. 2023. "Exploring Moral Algorithm Preferences in Autonomous Vehicle Dilemmas: An Empirical Study." Frontiers in Psychology 14: 1229245. https://doi.org/10.3389/fpsyg.2023.1229245

16. Waymo. 2020. Waymo Is Opening Its Fully Driverless Service to the General Public in Phoenix. Retrieved from: https://blog.waymo.com/2020/10/waymo-is-opening-its-fully-driverless.html.

17. Waymo. 2022. Redefine How You Move around San Francisco. Retrieved from: https://waymo.com/waymo-one-san-francisco/.

18. Waymo. 2023. Ride with Us in the City of Angels. Retrieved from: https://waymo.com/waymo-one-los-angeles/#main.

19. Waymo. 2026. The Future of Transportation Is Here. Retrieved from: https://waymo.com/rides/#rides-map.

20. Xing, Yang, Chen Lv, Dongpu Cao, and Peng Hang. 2021. "Toward Human-Vehicle Collaboration: Review and Perspectives on Human-Centered Collaborative Automated Driving." Transportation Research Part C: Emerging Technologies 128: 103199. https://doi.org/10.1016/j.trc.2021.103199

21. Xu, Tao, Andrei Dragomir, Xucheng Liu, et al. 2022. "An EEG Study of Human Trust in Autonomous Vehicles Based on Graphic Theoretical Analysis." Frontiers in Neuroinformatics 16. https://doi.org/10.3389/fninf.2022.907942

22. Zou, Bin, Zhenming Xiao, and Mutian Liu. 2020. "Driving Behavior Recognition Based on EEG Data from a Driver Taking over Experiment on a Simulated Autonomous Vehicle." Journal of Physics: Conference Series 1550 (4): 042046. https://doi.org/10.1088/1742-6596/1550/4/042046