Towards Context-Aware Modeling of Situation Awareness in Conditionally Automated Driving

Lilit Avetisyan

Industrial and Manufacturing Systems Engineering, University of Michigan-Dearborn

X. Jessie Yang

Industrial and Operations Engineering, University of Michigan-Ann Arbor

Feng Zhou

Industrial and Manufacturing Systems Engineering, University of Michigan-Dearborn

Corresponding author: Feng Zhou, 4901 Evergreen Road, Dearborn, MI 48128,

Email: fezhou@umich.edu

1   ABSTRACT

Maintaining adequate situation awareness (SA) is crucial for the safe operation of conditionally automated vehicles (AVs), which requires drivers to regain control during takeover (TOR) events. This study developed a predictive model for real-time assessment of driver SA using multimodal data (e.g., galvanic skin response, heart rate and eye tracking data, and driver characteristics) collected in a simulated driving environment. Sixty-seven participants experienced automated driving scenarios with TORs, with conditions varying in risk perception and the presence of automation errors. Using data from forty-four participants (twenty-three of those had invalid data) a LightGBM (Light Gradient Boosting Machine) model trained on the top 12 predictors identified by SHAP (SHapley Additive exPlanations) achieved promising performance with $RMSE = 0.89, MAE = 0.71$, and $Corr = 0.78$. These findings have implications towards context-aware modeling of SA in conditionally automated driving, paving the way for safer and more seamless driver-AV interactions.

Keywords: Situation awareness, real-time prediction, physiological measures, eye tracking, SHAP, autonomous vehicles.

2   Introduction

The rapid advancement of automated vehicle (AV) technologies holds the promise of transforming transportation. Yet, as vehicles progress through the levels of automation set by the Society of Automotive Engineers (SAE), they reach an intermediate phase known as conditionally automated driving – SAE Level 3 (SAE, 2021). In this phase, drivers must be ready to retake control in critical situations after receiving a takeover request (TOR) (Ayoub, Zhou, Bao, & Yang, 2019; Zhou, Yang, & De Winter, 2021; Zhou, Yang, & Zhang, 2019). Research consistently demonstrates that delayed and ineffective driver response during takeovers, especially when distracted by non-driving related tasks (NDRTs), substantially harm driving safety (Du et al., 2019; Du, Yang, & Zhou, 2020; Du, Zhou, et al., 2020a).

In conditionally automated vehicles, the maintenance of driver Situation Awareness (SA) – the accurate perception, comprehension, and projection of environmental elements (Endsley, 1995) – is paramount. The criticality of SA in ensuring timely and effective driver response during takeovers is well documented (Salmon, Stanton, Walker, & Green, 2006). However, driver engagement in NDRTs can lead to degradation of SA, leading to failures in takeovers (Endsley & Kiris, 1995; Körber, Baseler, & Bengler, 2018).

Risk perception, closely interwoven with SA, influences a driver's subjective assessment of potential threats, directly affecting trust in automation, vigilance levels, and takeover decisions (Hulse, Xie, & Galea, 2018; Seppelt & Lee, 2019). However, the existing literature focuses predominantly on the behavioral ramifications of risk perception and SA during manual driving, with less emphasis on their interplay within the scope of conditional automation where driver disengagement factors are distinct and heightened (Khastgir, Birrell, Dhadyalla, Sivencrona, & Jennings, 2017; Pop, Shrewsbury, & Durso, 2015).

Traditional SA measurement methods, such as freeze-probe techniques (e.g., SAGAT) or observer rating scales (e.g., SART) (de Winter, Eisma, Cabrall, Hancock, & Stanton, 2019; Durso et al., 1998), while reliable in certain contexts, are limited in their ability

to capture the continuous and dynamic changes of SA. Their intrusive nature and reliance on retrospective self-reports or observer judgments make them unsuitable for tracking the real-time evolution of SA in a rapidly changing driving environment. This knowledge gap highlights the need for developing unobtrusive, real-time methodologies to assess SA that utilize objective indicators, such as physiological signals and eye-tracking data.

Notably, advancements in machine learning offer promising avenues for the prediction and real-time monitoring of SA in automated driving, with studies indicating potential correlations between driver's SA levels and various physiological and behavioral markers (Du, Zhou, et al., 2020b; Zhou et al., 2021). Yet, research lacks a comprehensive analytical framework capable of integrating multimodal datasets to reliably predict SA, whilst considering individual differences and fluctuating driving conditions (Perello-March, Burns, Woodman, Elliott, & Birrell, 2021; Smith, Endsley, & Clark, 2023).

To bridge these gaps, the present study leverages a multimodal dataset, encompassing physiological responses and eye-tracking metrics, augmented by individual driver characteristics, to develop and validate a predictive model of SA tailored to automated driving scenarios. Specifically, a driving simulator experiment was conducted with 67 participants who experienced TOR events under varying risk perception and automation error conditions. Physiological data, i.e, GSR (Galvanic Skin Response), HR (Heart Rate), and HRV (Heart Rate Variability) were recorded using wearable sensors. Additionally, participants' eye movements were tracked to extract metrics such as fixation numbers, fixation duration, and dispersion across areas of interest like the center, left and right sides of road scene, NDRT display and odometer. Self-reported SA ratings were collected every 30 seconds during the drives as a ground truth.

Further, this study employs Light Gradient Boosting Machine (LightGBM) (Ke et al., 2017) and SHapley Additive exPlanations (SHAP) (Lundberg et al., 2020; Lundberg & Lee, 2017a) to not only predict SA but also unpack the contribution of each feature to the model's decisions. In doing so, it endeavors to contribute to the design of

context-aware SA monitoring systems that enhance the safety and efficiency of driver-AV symbiosis.

The contributions of this paper are thus summarized as:

- Development of a non-intrusive LightGBM model that leverages multimodal sensor data for real-time SA assessment in conditionally automated driving.

- Identification and analysis of key physiological and behavioral predictors for SA using SHAP values.

- Exploration of the interplay between risk perception, driver characteristics, and SA in the context of conditional automated driving.

- Demonstration of the practicality and effectiveness of applying machine learning for SA prediction to foster improved driver-AV interaction.

3   Related Work

3.1   Importance of SA in Conditionally Automated Driving

SA, 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" (Endsley, 1988), plays a critical role in ensuring safe and efficient interactions between humans and AVs. Drivers need to maintain sufficient SA to understand the current driving context, anticipate potential hazards, and respond appropriately to changing situations, especially during the takeover transition process. However, the unique characteristics of conditionally automated driving environments present new challenges to SA, including mode confusion and automation complacency. Mode confusion: Drivers may struggle to understand the current operational mode (manual vs. automated) and their respective roles and responsibilities. This uncertainty can lead to inappropriate actions and delayed responses in critical situations (Merat & Jamson, 2009). Automation complacency: Overreliance on automation can lead to decreased vigilance and reduced monitoring of the driving environment, resulting in impaired SA and increased accident risk (Thill, Hemeren, & Nilsson, 2014). Delayed

takeover transitions: Smooth and timely transfer of control between the driver and the automated system is crucial for maintaining SA and avoiding accidents. Difficulty in resuming control can lead to confusion and errors, especially in complex or unexpected situations (Zhou et al., 2022).

3.2   The Impact of Risk Perception on SA

Risk perception, the subjective judgment of the likelihood and severity of potential hazards, plays a crucial role in influencing SA. Drivers who perceive a higher level of risk tend to be more vigilant and attentive to their surroundings, leading to enhanced SA. Conversely, low risk perception can lead to complacency and reduced awareness of potential dangers, ultimately impairing SA and increasing the likelihood of accidents. Several factors influence risk perception in automated driving:

Automation error: Drivers' trust in the automated system's capabilities significantly impacts their risk perception. If the system is perceived as reliable, drivers may be less vigilant and have lower SA. Conversely, experiences with system failures or unexpected behavior can increase perceived risk and enhance SA (Merat & Jamson, 2009; Thill et al., 2014). System transparency: Understanding how the automated system works and its limitations is crucial for accurate risk assessment. A lack of transparency can lead to uncertainty and distrust, negatively affecting risk perception and SA (Manchon, Beaufort, Bueno, & Navarro, 2022; Schwarz, Gaspar, & Brown, 2019). Hazard levels: The presence of potential hazards in the driving environment naturally influences risk perception. Drivers are more likely to be aware of and attentive to situations with higher perceived risk, leading to increased SA (Niu et al., 2022). Individual differences: Drivers' age, experience, personality traits, and cultural background can influence their risk perception and subsequently, their SA. For example, older drivers may have lower risk tolerance and exhibit higher levels of SA compared to younger drivers (M. Li, Holthausen, Stuck, & Walker, 2019).

3.3   Measuring SA: From Subjective Assessments to Objective Measures

Accurately assessing SA in conditionally AVs is essential for understanding driver behavior and developing effective safety interventions. Traditional SA measurement methods primarily rely on subjective assessments. Situation Awareness Global Assessment Technique (SAGAT) is widely used method that involves freezing the driving scenario and asking participants a series of questions about the current situation to assess their level of awareness (Endsley, 1995). Situation Awareness Rating Technique (SART) involves trained observers rating an individual's SA based on their performance in a simulated driving task, considering factors such as scanning behavior, response times, and decision-making (Taylor, 2017). Situation Present Assessment Method (SPAM) method combines subjective self-assessments via questionnaires with objective performance measures, such as reaction times and driving errors, to provide a more comprehensive evaluation of SA (Durso et al., 1998). While these subjective measures offer valuable insights, they suffer from limitations such as susceptibility to biases, reliance on memory and self-perception, and inability to capture the dynamic fluctuations of SA over time (Salmon et al., 2006). To overcome these limitations, researchers have explored the use of objective and continuous measures based on physiological and eye-tracking data: Electroencephalography (EEG): By measuring brainwave activity, EEG can provide insights into cognitive workload, attentional focus, and mental fatigue, all of which are related to SA (Fernandez Rojas et al., 2019; Yeo et al., 2017). Heart Rate Variability (HRV): Variations in heart rate reflect changes in the autonomic nervous system, which is influenced by cognitive and emotional states relevant to SA (Perello-March et al., 2021). Galvanic Skin Response (GSR): This measure reflects changes in skin conductance, which is associated with arousal and emotional responses that can indicate changes in SA (Smith et al., 2023). Eye-tracking metrics: By monitoring eye movements, researchers can analyze fixation duration, saccade patterns, and scan paths to understand where drivers are focusing their visual attention and how efficiently they are processing information (de Winter et al., 2019). Studies have demonstrated that these objective measures can provide more accurate

and reliable assessments of SA compared to subjective methods, particularly in dynamic and complex driving environments. For a comprehensive review of physiological measurements of situation awareness, please refer to (Zhang et al., 2023).

3.4   Predicting SA: Towards a Multimodal and Individualized Approach

Recent research has explored the use of machine learning and context-aware models to predict SA in real-time. These models leverage various data sources, including physiological measures, eye-tracking data, vehicle sensor data, and subjective ratings, to estimate drivers' level of awareness and predict potential lapses in attention. Some notable approaches include: By analyzing physiological and behavioral data, deep learning algorithms can effectively predict driver SA with high accuracy. Li et al. (2022) proposed a deep learning model that achieved promising results in predicting driver SA during conditionally automated driving scenarios. Eye-tracking based models: Eye movements provide valuable insights into drivers' visual attention and information processing. Zhou et al. (2021) developed a machine learning model using eye-tracking data to predict SA levels in real-time. Functional near-infrared spectroscopy (fNIRS) allows for direct measurement of brain activity during driving tasks. Unni et al. (2017) used fNIRS to identify distinct neural signatures associated with different levels of SA. By incorporating information about the driving environment, traffic conditions, and driver behavior, context-aware models can provide more accurate and personalized predictions of SA. Zhang et al. (2018) demonstrated the effectiveness of incorporating contextual information in predicting driver SA. While these advancements show promise, most existing research focuses on single modalities or specific contexts. A more comprehensive approach is needed that integrates various data sources and considers individual differences to develop robust and personalized SA prediction models.

Our research aims to address the existing gaps in SA prediction by developing a multimodal and individualized approach. We will collect and analyze physiological measures, eye-tracking data, and individual characteristics to build a comprehensive model that predicts SA in conditionally automated driving scenarios. This research

holds significant potential for developing real-time SA monitoring systems that can provide timely interventions and improve driver safety and performance in AVs. Our study will contribute to the growing body of knowledge on SA in automated driving and pave the way for more personalized and effective driver support systems in the future.
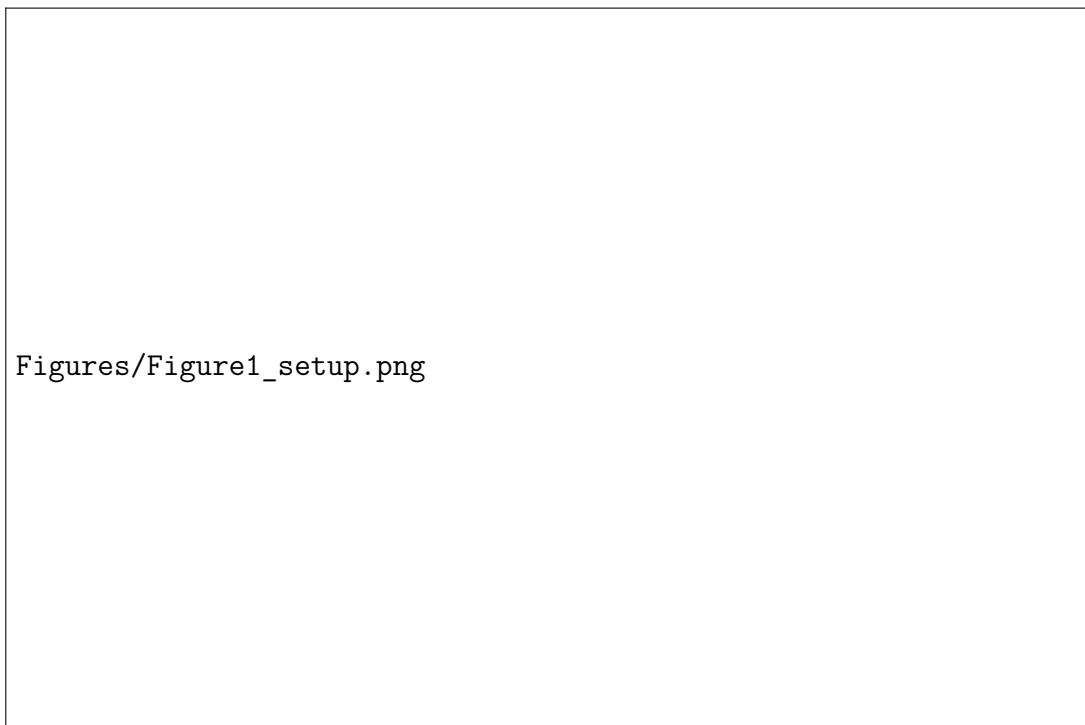
4   Methodology

4.1   Participants

A total of 67 people (30 females: mean age = 28.3, SD = 11.5; and 37 males: mean age = 25.9, SD = 12.3) participated in this study. Due to malfunction of physiological sensors and the driving simulator, 23 participants were excluded, and data from the remaining 44 participants (21 females: mean age = 29.09, SD = 12.35; and 23 males: mean age = 25.75, SD = 10.89) were used for further analysis. All the participants had a valid driver's license (mean years of experience = 9.1, SD = 11.7). Participants received $30 in compensation for about 75 min of participation. The study was approved by the Institutional Review Board at the University of Michigan.

4.2   Apparatus and stimuli

This research utilized a desktop-based driving simulator by Realtime Technologies Inc. (RTI, Michigan, USA) to gather experimental data. The simulator system included an array of three LCD monitors, a Logitech G29 driving kit, one tablet for engagement in NDRTs, and one phone, positioned to the participant's right side, for recording SA assessments (see Fig. 1). The tablet and phone were moved to the left side for left-handed participants upon request. For the NDRT, a specially engineered Tetris game was developed using the PyGame library within the Python programming environment. The game's flow allowed participants to engage with the game tiles upon NDRT initiation, it automatically paused when TORs were triggered, enabling a seamless resumption from the previous state during the next NDRT request. The driving simulation system was set to emulate a vehicle with conditionally automated driving (SAE Level 3) capabilities (SAE, 2021). To engage in the automated

Figure 1. The experimental setup. Participants were seated in front of a three-monitor display, using a steering wheel and pedals to interact with the simulated driving environment. Eye-tracking glasses recorded gaze patterns, while a GSR device worn on the participant's left hand measured physiological arousal. Additionally, a tablet situated to the participant's right administered the NDRT, and a nearby smartphone recorded self-reported SA ratings.

drive mode, participants were instructed to press a red button positioned on the steering wheel. Upon the engagement of this mode, participants received a confirmation, an audio "Automated mode engaged" prompt, and the mode indicator on the odometer turned white. Then, the AV continued navigating a pre-defined route at a steady speed of 35 mph. While experiencing automated driving, participants were requested to start the NDRT (i.e., the Tetris game on a tablet) upon receiving the "Please start the secondary task" audio prompt. When a TOR – a "Takeover" audio request was initiated, participants were alerted to disengage the automated mode and take manual control of the vehicle. If a participant was unable to resume vehicle control within seven seconds, an "Emergency Stop" audio alarm was activated, and the AV was triggered to stop immediately.

The self-reported SA assessment was conducted through a single-item questionnaire prompt (see Fig. 2) developed on the Qualtrics platform (Provo, UT,

www.qualtrics.com), and administered via a mobile phone. The simulation also recorded physiological responses. To capture the details of visual attention, the Pupil Core eye-tracker headset with a frequency of 200 Hz from Pupil Lab (MA, USA) was used, Concurrently, GSR and HR (via photoplethysmography or PPG) were recorded at 128 Hz using the iMotions platform with the Shimmer3 GSR+ Unit (Shimmer, MA, USA). To ensure precise synchronization of time across all devices, including the iMotions, Pupil Core, driving simulator, and Tetris game, the time delays were recorded at the moment of the experiment's initiation.

4.3 Experimental design

The study employed a 2x2 mixed design experiment where the between-subjects factor was the risk condition (high-risk vs. low-risk), and the within-subjects factor was the automation errors (error vs. no error). Participants were randomly assigned to one of the risk conditions: 1) a high-risk condition, where participants were presented with negative information about AV performance through videos showcasing its malfunctions coupled with an environment simulation of driving in foggy weather. The video analyzes an AV autopilot failure that led to a collision with a stationary, overturned truck on a highway, exploring potential contributing factors, supported by similar incidents, such as difficulties in detecting and responding to stationary objects, environmental conditions affecting performance, and unsafe accelerator inputs overriding the autopilot system. (Link: high-risk video); 2) a low-risk condition, where a sunny weather driving simulation was preceded by a video demonstrating the AV's ability to anticipate safety-critical hazards that are challenging for human drivers to detect. The pre-simulation video displayed real-life dashcam footage, demonstrating the AV's successful detection and response to potential dangers, such as merging vehicles, abrupt stops, swerving trucks, animals on the road, unsafe lane changes, and adverse road conditions (Link: low-risk video). These conditions were established using video examples derived from authentic environments.

Each participant experienced two drives with varying automation errors: 1) No Error:
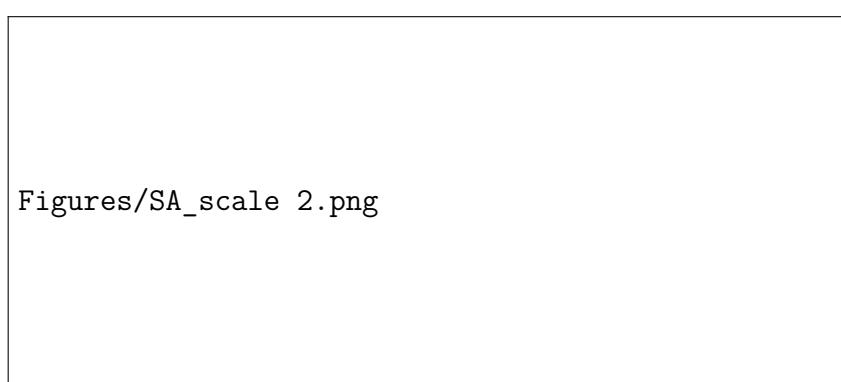
The AV issued four accurate TORs without technical faults, 2) Error: Participants experienced two accurate TORs (first and fourth) and two erroneous TORs (second: false alarm, third: miss). Standard road scenarios were used to trigger accurate TORs or simulate errors (e.g., pedestrians crossing, construction zones, accidents)(see Fig. 3). Previous research suggested both false alarms and misses can decrease user trust (Ayoub, Avetisian, Yang, & Zhou, 2023). The sequence of drives was counterbalanced across participants.

4.4 Experimental procedure

At the onset of the experiment, participants were briefed on the equipment and were provided with instructions regarding the tasks they would be performing. They were informed about the capabilities and limitations of Level 3 AVs, with a specific focus on the necessity for vigilance and readiness to take control whenever a TOR was issued. Following the briefing, participants underwent the device setup process. This involved attaching GSR electrodes to the participants' left palm and recording the baseline, fixing a PPG sensor to their left earlobe, and calibrating the eye-tracking device. To ensure a consistent understanding of SA levels and automation error types, participants engaged in several information sessions and practical demonstrations. First, they received a detailed explanation of SA levels used in the study, ranging from being unaware of the driving conditions (level 0) to predicting other road users' actions (level 3). Next, participants were introduced to the single-item SA survey presented on a mobile phone, which they would use to record their SA levels every 30 seconds during the experimental drives. The survey prompted them to rate their current SA on a 0-3 scale,aligning with the previously explained SA levels, with the instruction "Please indicate your situation awareness" (see Fig. 2). This frequent assessment allowed for continuous and unobtrusive monitoring of SA changes throughout the drives. Furthermore, participants were explicitly informed about the potential for encountering different automation error types during the automated driving phases, including errorless performance, misses (automation failure to detect hazards), and false alarms

(unnecessary initiated TORs). To solidify their understanding, participants engaged in a practice driving session, and experienced situations representing automation errors and different SA level needs. They practiced providing self-assessments of their SA level using the single-item survey and received feedback from the experimenter to clarify any misunderstandings and ensure accurate differentiation between SA levels and recognition of AV behaviors. Following these preparatory activities, participants completed an online survey providing demographic information and were presented with risk-manipulating video information, specific to their assigned risk condition. Finally, participants proceeded to the actual driving sessions, each lasting approximately 15 minutes. After each session, they completed surveys asking to evaluate their trust (Holthausen, Wintersberger, Walker, & Riener, 2020), emotional responses (Jensen et al., 2020), and perceived risk (Zhang et al., 2019) specifically related to their experience during that particular driving session. These measures were all evaluated using 7-point Likert scales.

The total duration of the experiment was approximately 75 minutes. Note the results of emotional responses were not reported in this paper as our focus here is SA.


Figure 2. SA level assessment prompt: Participants were instructed to report their situational awareness level every 30 seconds. To report, they adjusted the sliding bar to reflect their current SA level, or left it unchanged if their SA remained consistent with their previous submission.
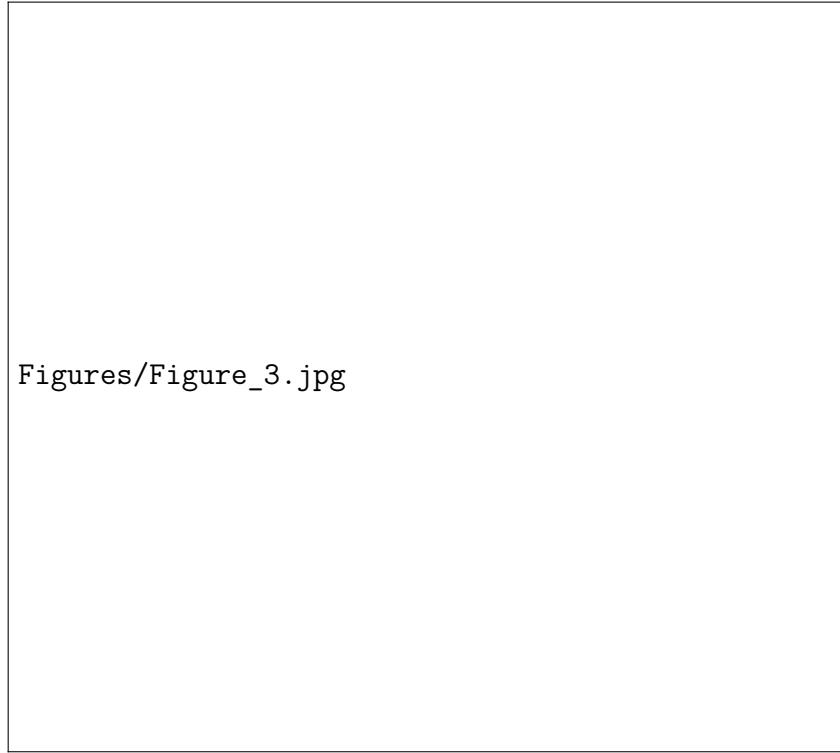
Figure 3. Takeover events in urban areas (a) pedestrians crossing ahead (b) bus sudden stop ahead (c) construction zone ahead (d) police vehicle on shoulder.

5 Predictive SA Model

5.1 Data Preprocessing

The study utilized a multimodal dataset comprised of physiological signals (GSR, HR, HRV), eye-tracking data, self-reported SA assessments, and demographic information. Here's the data preparation process:

GSR Processing: The GSR signal was decomposed into tonic and phasic components using the Neurokit2 package (Makowski et al., 2021). The phasic component, known for its sensitivity to rapid changes, was used for analysis.

HRV Calculation: The HRV was calculated from collected IBI (inter-beat interval) using the Root Mean Square of Successive Differences (RMSSD) method (see Eq. 1):

$$RMSSD = \sqrt{\sum_{i=1}^{N} \left(RR\_interval_i - RR\_interval_{i+1} \right)^2}, \tag{1}$$

where N is the number of heartbeats and the "RR interval" is the distance (in

milliseconds) between two consecutive successful heartbeats.

Eye-Tracking Metrics: Relevant eye-tracking metrics were extracted, including fixation numbers, fixation duration (in milliseconds), and dispersion (in degrees). Fixation numbers were identified using the I-DT algorithm (Salvucci & Goldberg, 2000) (maximum dispersion: 1 degree, maximum duration: 200ms).

Synchronization and Alignment: Data from all sources (iMotions, Pupil Core, Qualtrics) were synchronized using timestamps across the recorded platform delays. A 30-second sliding window was applied to calculate average GSR, HR, and eye-tracking metrics within each window. Self-reported SA ratings were linked to the physiological and eye-tracking data based on timestamps.

Integration: The final dataset integrated physiological, eye-tracking, and demographic data (age, gender, AV knowledge level), resulting in the 21 features outlined in Table 1.

5.2   LightGBM model

This research aims to build a real-time predictive model for driver's SA using a multimodal dataset of physiological signals (i.e., GSR, HR, HRV), eye-tracking data, alongside self-reported assessments and demographic variables. Due to the continuous and hierarchical structure of SA, which encompasses levels that exist on a spectrum rather than in discrete categories, a regression approach was adopted to predict SA and analyze its dynamic changes during automated driving. After evaluating several regressor algorithms within the Azure ML environment, including Random Forest, Decision Trees, XGBoost, LightGBM, and Neural Networks, LightGBM was selected for this study due to its advantageous performance based on RMSE and MAE metrics (see Table 2).

LightGBM is widely used in handling large datasets and high-dimensional features for regression tasks. It employs a leaf-wise growth strategy for decision trees, which can lead to better accuracy with less computation when compared to depth-wise growth strategies used by other algorithms. It is also well-suited for scenarios where speed and memory usage are critical, such as AVs, without compromising on model performance.

Moreover, LightGBM supports advanced techniques like Gradient-based One-Side Sampling (GOSS) to reduce memory usage and improve training speed, and Exclusive Feature Bundling (EFB) to reduce the number of features and improve the efficiency of the model. Another advantage of LightGBM is that it effectively handles categorical or ordinal features directly, e.g. AV knowledge level in our dataset. During training, LightGBM considers these values as categorical features and looks for the best splits based on the categorical nature of the data. In this study, LightGBM regressor was employed to predict driver's SA. First, the model's performance was optimized using a grid search for hyperparameter tuning, aiming to minimize the root mean square error (RMSE) and mean absolute error (MAE) as performance metrics. Based on the tuning results, the following values were set for the model parameters: 'objective': 'regression', 'metric': {'mae', 'rmse'}, 'learning_rate': 0.05, 'min_data_in_leaf': 20, 'num_leaves': 50, 'early_stopping_rounds': 100. Next, the model was trained and validated using the dataset of 21 features and 1634 samples, and using a 10-fold cross-validation methodology.

Table 1

Prediction Model Features. The asterisk (*) denotes the feature's importance in the prediction model.

Features	Unit	Description
1. age *	years	Participant's age
2. avKnowledge *	-	Participant's knowledge level about AVs
3. gender *	-	Participant's gender
4. mean_gsr *	μS	Average galvanic skin response in phasic phase
5. mean_HR *	bpm	Average number of heartbeats
6. mean_HRV *	ms	Average of the variation in the time interval between heartbeats
7. number_of_fixations_center *	-	Number of fixations on the center screen
8. number_of_fixations_game *	-	Number of fixations on the game display
9. number_of_fixations_left	-	Number of fixations on the left screen
10. number_of_fixations_right	-	Number of fixations on the right screen
11. number_of_fixations_odometer	-	Number of fixations on the odometer
12. mean_dispersion_center *	degree	Average distance between all gaze locations during a fixation on center screen
13. mean_dispersion_game *	degree	Average distance between all gaze locations during a fixation on game screen
14. mean_dispersion_left	degree	Average distance between all gaze locations during a fixation on left screen
15. mean_dispersion_right	degree	Average distance between all gaze locations during a fixation on right screen
16. mean_dispersion_odometer	degree	Average distance between all gaze locations during a fixation on odometer screen
17. mean_duration_center *	ms	Average duration of fixations on the center screen
18. mean_duration_game *	ms	Average duration of fixations on the game screen
19. mean_duration_left	ms	Average duration of fixations on the left screen
20. mean_duration_right	ms	Average duration of fixations on the right screen
21. mean_duration_odometer	ms	Average duration of fixations on the odometer screen

5.3   SHAP Explainer

To reveal the contribution of each feature in the predictive model of SA, the SHAP (Lundberg & Lee, 2017b) approach was used. SHAP values provide a consistent and locally accurate method to attribute the effect of each feature in a prediction task, based on the foundational principles of Shapley values from cooperative game theory (Kuhn & Tucker, 1953). It ensures that each feature receives an importance weight by averaging over all possible permutations of feature orderings while considering the interaction effects between features. Moreover, SHAP model provides a detailed explanation for local/individual predictions, as well as aggregate SHAP values across multiple instances, offering a broader view of feature importance and model behavior. As the SA prediction model training process used a 10-fold cross-validation, the SHAP values were calculated ten times, once for each fold. The final impact of each feature was then determined by the average of these ten sets of SHAP values, providing a measure of the average contribution over the entire cross-validation process (Ayoub et al., 2023; Ayoub, Yang, & Zhou, 2021).

6   Results

6.1   Model performance

The LightGBM model's performance in predicting SA is presented in Table 2. To analyze how individual features contribute to predictions, SHAP values were calculated and visualized in a SHAP summary plot (Fig. 4). This reflects both the importance and the effects of each feature. The y-axis positioning is determined by the feature importance, ranging from the most to the least significant. The x-axis is determined by the SHAP value, where every point on the plot represents a feature's SHAP value for a given instance. The color scale, ranging from blue (low) to red (high), indicates the magnitude and direction of a feature's impact on the predicted SA score, and the overlapping points, jittered in the y-axis direction, describe the distribution of the SHAP values per feature. To further optimize the model's performance, the model's behavior was tested by adding the features incrementally according to their SHAP
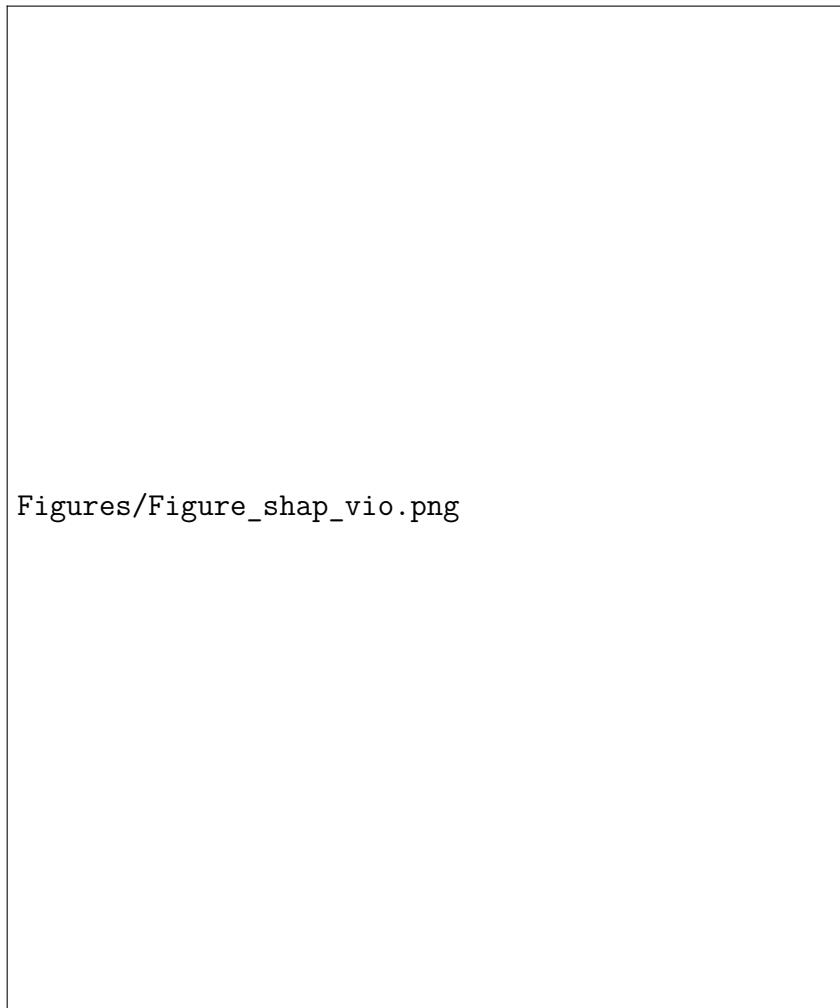
Figure 4. SHAP summary plot. The x-axis shows the feature's influence on SA. The y-axis shows the importance ranking of the features.

importance rankings. The impact of each addition on the model's accuracy is illustrated in Fig. 5, which shows the variations in performance metrics (i.e., RMSE, MAE, and Corr) with the inclusion of more predictors. It was observed that the LightGBM model resulted in improved performance with a subset of top 12 features rather than the entire feature set. Following this insight, the model was retrained with these 12 features where the updated model had the best performance (see Table 2 under the 'Selected features' row).

According to SHAP and model performance, the following features had the most impact on the model and are listed in descending order: age, avKnowledge, mean_gsr, mean_HR, mean_HRV, number_of_fixations_game, number_of_fixations_center, mean_dispersion_game, gender, mean_duration_game, mean_duration_center,
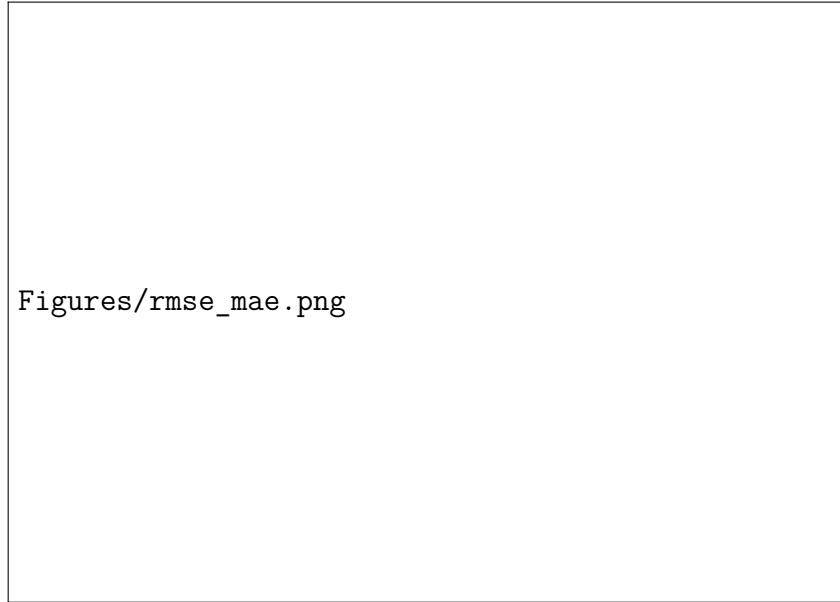
Figure 5. Iterative improvement of LightGBM model performance across 21 iterations, where features were incrementally added according to their importance ranking as determined by SHAP values.

mean_dispersion_center. However, the summary plot shows only the global view of how feature values influenced predictions.

Table 2
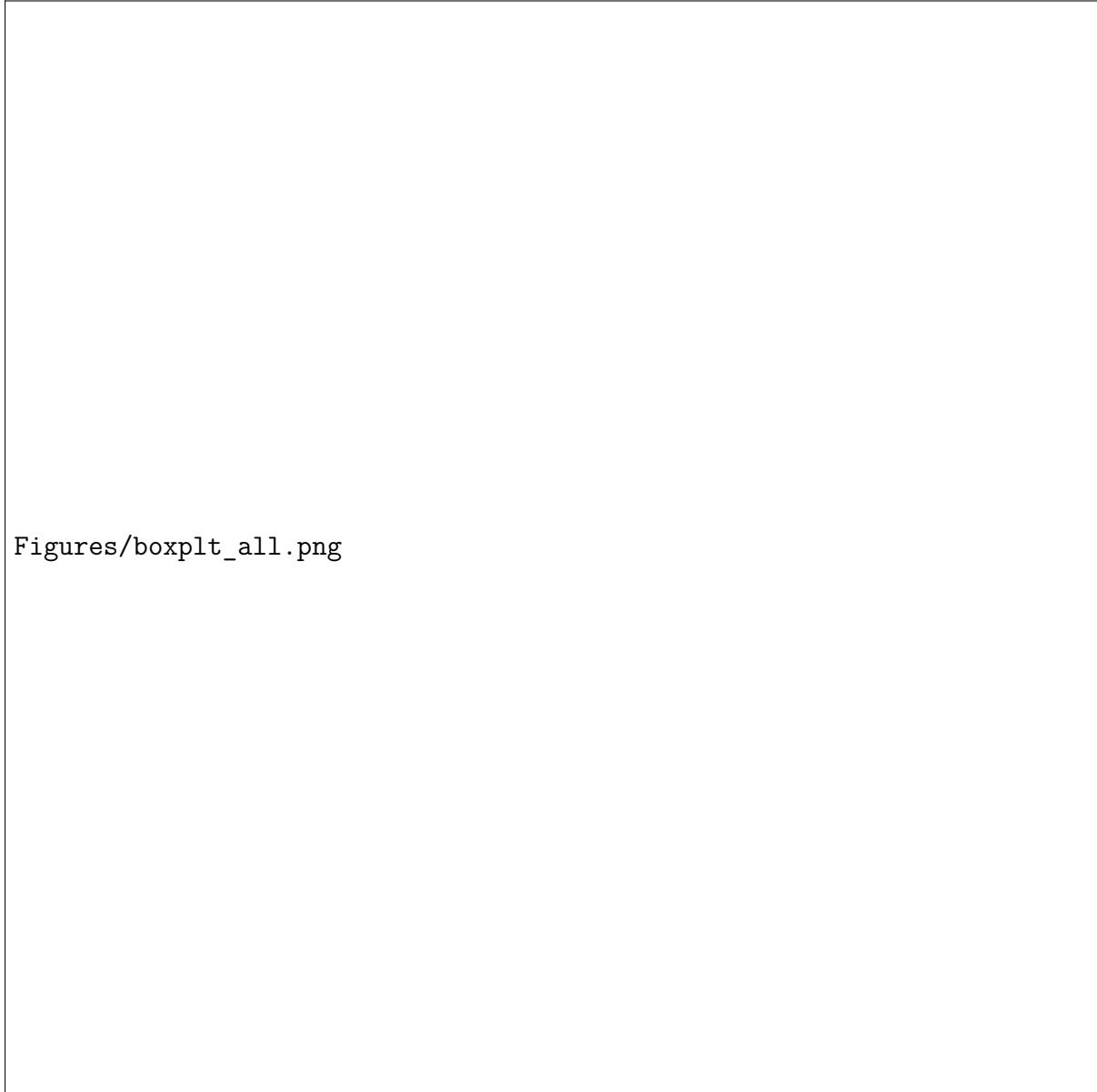
Performance of tested regressor models.

Model	RMSE	MAE	Corr
Neural Network	1.02	0.87	0.80
XGBoost	0.95	0.78	0.78
Random Forest	0.94	0.77	0.78
Decision Tree	0.94	0.76	0.77
LightGBM (All features)	0.90	0.72	0.77
LightGBM (Selected Features)	0.89	0.71	0.78

6.2   Feature Effect in Predicting SA

To explore the impact of an individual feature on the SA predictions made by the model, for the top 12 features, the SHAP value was charted against the feature's actual values. This relationship is shown in Fig. 6, where the feature values were segmented into several groups according to their distribution (the physiological features were

limited to 20 segments). This segmentation allowed the examination of the feature's effect on SA. Spearman's rank correlation coefficient for each feature was calculated to understand the strength of the relationship.


Figures/boxplt_all.png

Figure 6. The effect of important features on predicted SA value. The x-axis represents the value of the feature, and the y-axis represents the SHAP value associated with that feature. Positive SHAP values indicate that the feature pushes the prediction higher, while negative values indicate the opposite.

Age: The most important feature was age in Fig. 4 with a significant negative correlation ($\rho = -0.294, p < 0.001$). A complex relationship with SA was observed. Younger adults (18-38) showed minimal influence of age on SA. A strong negative effect was found for middle-aged adults (39-62) (i.e., those in this age range had significantly

lower predicted SA), while older adults (62+) displayed a positive effect. However, the dataset's skewed age distribution warrants caution in interpretation.

AV Knowledge: Participants with higher AV knowledge tended to have lower SA predictions ($\rho = -0.665, p < 0.001$).

Physiological Signals: GSR: The mean_gsr was positively linked to SA ($\rho = 0.463, p < 0.001$), with higher values suggesting an increase in SA, though the relationship was not strictly linear. Heart Rate: mean_HR showed a positive relationship with SA ($\rho = 0.670, p < 0.001$), indicating that elevated HR was associated with higher levels of SA. HRV: Although the correlation was significant ($\rho = 0.128, p < 0.001$), the effect size suggests a very weak association with no overall trend.

Eye-Tracking Metrics: Fixations (Game): A significant positive correlation was found between the number_of_fixations_game and SA ($\rho = 0.724, p < 0.001$), suggesting that an increased frequency of fixations correlates with enhanced SA. Fixations (Center): There was a notable pattern characterized by a decrease in SA with a lower frequency of fixations at the center($\rho = -0.302, p < 0.001$). Dispersion (Game): The mean_dispersion_game seemed to be negatively correlated with SA ($\rho = -0.233, p < 0.001$), indicating that greater distances between fixation points on game surface were associated with lower levels of SA. Dispersion (Center): The mean_dispersion_center ($\rho = 0.310, p < 0.001$) seemed to be positively correlated with higher dispersion on center leading to increase of SA. Fixation durations: Mean duration of fixations on both the game and center surfaces showed weaker correlations, with less clear trends.

Gender: The data showed that female participants tend to have higher SA than males ($\rho = -0.670, p < 0.001$).

6.3 SA, Trust and Perceived Risk Across Conditions

A mixed two-way analysis of variance (ANOVA) was used to analyze the effects of risk perception and automation error on participants' SA, trust, risk, and physiological

responses. The ANOVA showed a significant main effect of automation error $(F(1, 56) = 5.313, p = 0.025, \eta_p^2 = 0.087)$ and a marginal main effect of risk condition $(F(1, 56) = 3.438, p = 0.069, \eta_p^2 = 0.058)$ on SA. Within the high-risk group, participants reported a significantly higher level of SA $(p = 0.018)$ during the drive with automation error compared to the low-risk group. The Fig. 7 visualizes the dynamic SA reported by participants in the experiment, showing how perceived risk and automation error influence driver awareness. SA initially declined in both conditions as drivers adapted to automation, then oscillated as drivers experienced TORs which temporarily boosted SA. When drivers re-engaged in NDRT, SA decreased again. Participants in the high-risk group consistently showed higher SA, which confirmed that emphasizing potential hazards and limitations of automated driving (i.e. risk perception manipulation) leads drivers to maintain a higher level of awareness and vigilance. On the other hand, automation errors (i.e., false alarms and missed TORs) further elevated SA by increasing drivers' distrust and alertness. In terms of trust, a significant main effect of automation error was found $(F(1, 56) = 13.700, p < 0.001, \eta_p^2 = 0.181)$. For risk perception, although no significant differences were found between the two conditions, it successfully elicited different levels of self-reported SA.

6.4 Objective Responses Across Conditions

The effect of risk and automation error was also investigated for physiological features using a two-way mixed ANOVA. There was a significant interaction effect on mean_HR $(F = 7.348, p = 0.012)$, with a large effect size $(\eta_p^2 = 0.242)$. For mean_duration_center, there were significant main effects of both risk $(F(1, 42) = 5.34, p = 0.03, \eta_p^2 = 0.188)$ and automation error $(F = 9.077, p = 0.006, \eta_p^2 = 0.283)$, as well as a significant interaction effect $(F(1, 42) = 5.089, p = 0.034, \eta_p^2 = 0.181)$. The pairwise t-test showed that drive with automation errors resulted in longer fixation duration in the center compared to drive where no error was experienced. Additionally, the high risk condition led to longer center fixation duration than the low risk condition. Mean_dispersion_center showed
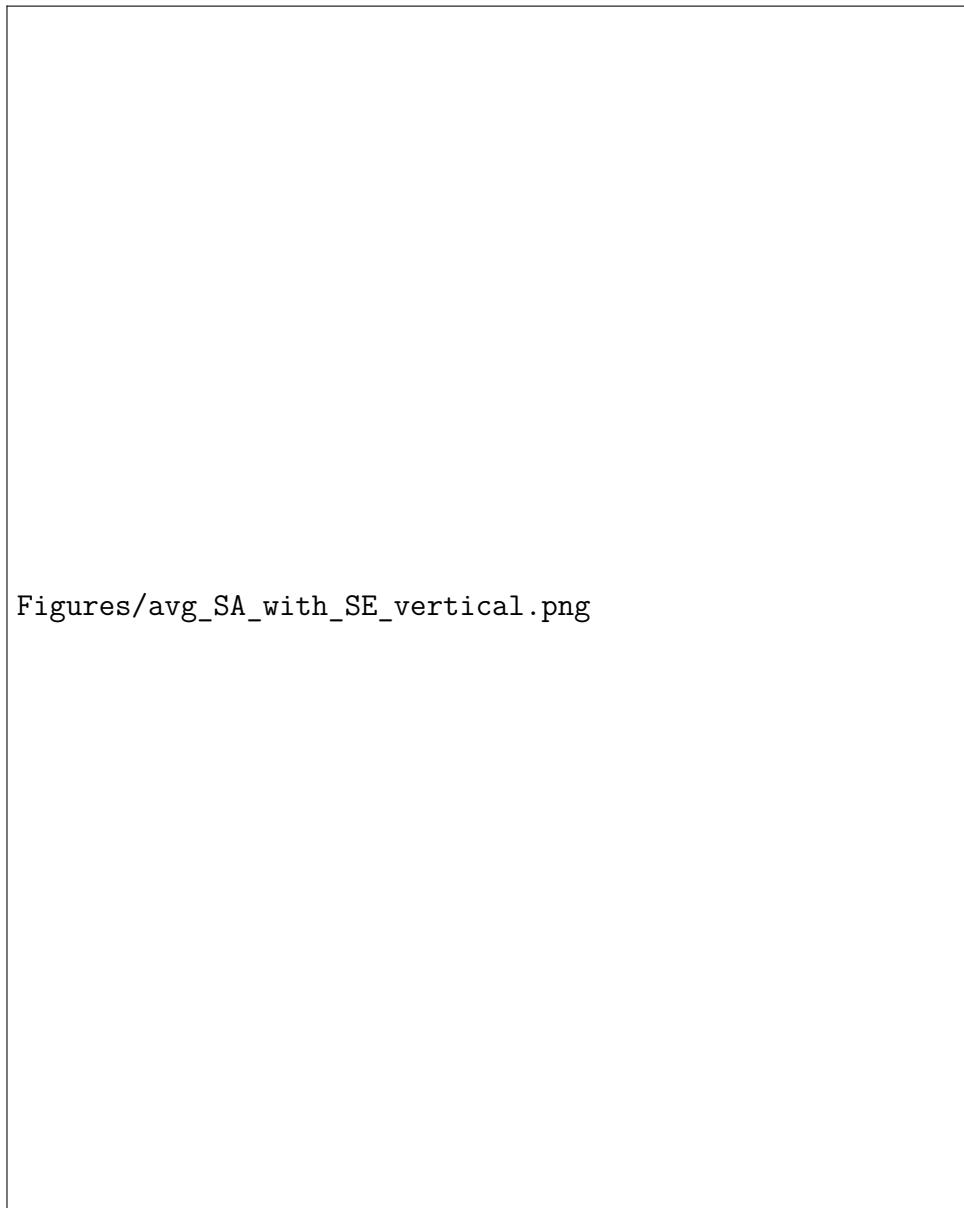
Figure 7. Dynamic changes in self-reported SA across different risk conditions and automation error scenarios. The x-axis represents the sequential number of each 30-second interval during which SA was reported. The y-axis represent the average SA. Error bars represent standard error.

significant main effects of risk ($F = 5.988, p = 0.022, \eta_p^2 = 0.207$) and automation error ($F = 4.657, p = 0.042, \eta_p^2 = 0.168$), where automation error increased dispersion in the center area, similarly, the high risk environment increased dispersion compared to low risk. The automation error had a significant main effect on number_of_fixations_center ($F = 10.786, p = 0.003, \eta_p^2 = 0.319$) where fixation number was higher when AV had errors. For the number_of_fixations_game automation error showed significant main effect as well($F = 8.423, p = 0.008, \eta_p^2 = 0.268$) where participants had more fixations on

game display during the drives without automation error. Risk significantly impacted on mean_dispersion_left ($F = 5.516, p = 0.028, \eta_p^2 = 0.193$) and mean_dispersion_right ($F = 4.41, p = 0.047, \eta_p^2 = 0.161$), with high risk leading to greater dispersion compared to low risk. Finally, automation error had significant main effects on mean_duration_odometer ($F = 4.698, p = 0.041, \eta_p^2 = 0.17$) and mean_dispersion_odometer ($F = 6.817, p = 0.016, \eta_p^2 = 0.229$), with higher duration and dispersion when automation had errors.
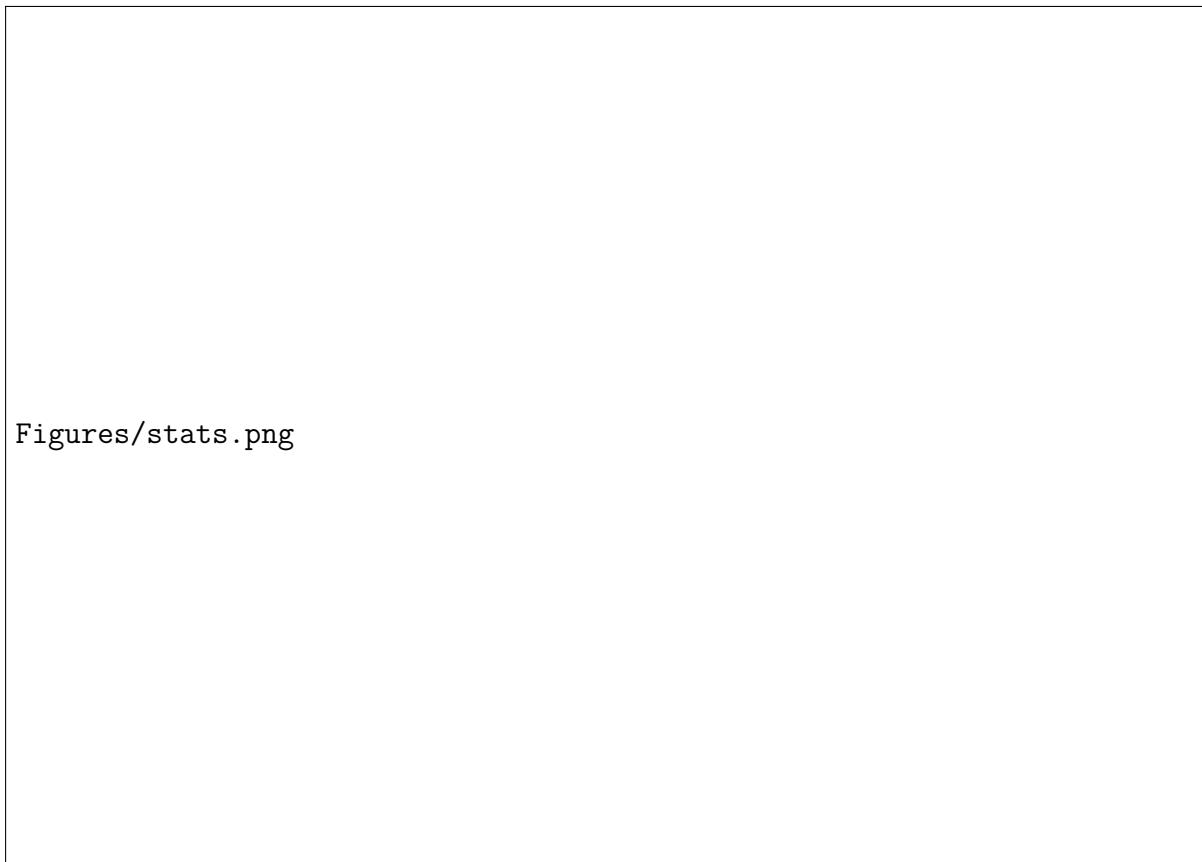
Figure 8. Main effects on features across tested conditions.

7 Discussion

7.1 Predicting SA with machine learning model

In our investigation, the LightGBM machine learning model has demonstrated promising capabilities in estimating SA from a diverse array of signals. By incorporating a multimodal dataset—including physiological signals, eye tracking metrics, and

demographic information—the model was able to predict SA levels with RMSE of 0.90, MAE of 0.72, and a correlation coefficient of 0.77 with self-reported SA measures.

To contextualize these results, we employed the SHAP (SHapley Additive exPlanations) framework to interpret the predictive model. The SHAP summary plot not only illuminated the overarching importance of each feature but also revealed the nature of their effects on SA prediction. It is noteworthy that we identified certain features whose contribution appeared marginal, potentially serving as noise that detracted from the model accuracy (Ayoub et al., 2023). When considering a refined feature set with the top 12 variables as indicated by their SHAP values, the model's performance was marginally enhanced, showing an RMSE of 0.89, an improved MAE of 0.71, and a boosted correlation of 0.78.

Among the principal features influencing the model's predictions were demographic aspects such as age, gender and experience with automated vehicles. This seemed to be consistent with previous findings that demographic variables could reflect diverse cognitive abilities and confidence levels in tasks, thereby impacting SA in an automated context (Kintz, Banerjee, Zhang, Anderson, & Clark, 2023; S. Li, Blythe, Guo, & Namdeo, 2018).

Physiological signals, including GSR, HR, and HRV, were also important. Mean_gsr was positively correlated to SA, possibly because a higher level of GSR was related to a high level of arousal and alertness (Zhou, Qu, Helander, & Jiao, 2011; Zhou, Qu, Jiao, & Helander, 2014), which further led to a higher level of arousal. Mean_HR was also positively correlated with SA, which might be a pertinent indicator of participants' stress levels reacting to potential risks and errors of the AV, which was integral for the participants to pay more attention to the driving situations (Perello-March, Burns, Birrell, Woodman, & Elliott, 2022). Previous studies also found that a higher value of HRV was associated with better attentional maintenance and flexibility, which might have contributed to its positive correlation with SA (Siennicka et al., 2019).

We found a positive correlation between the number of fixations on the game display and SA and a negative correlation between the number of fixations on the center area

(presumably the road ahead) and SA. Such findings are somewhat counterintuitive. Previous studies showed that increased visual attention towards the NDRTs (i.e., game) suggested a lack of focus on the driving environment, potentially leading to lower SA and typically more frequent fixations on the road were associated with better SA (Liang, Yang, Prakah-Asante, et al., 2021; Zhou et al., 2021). However, such findings could be influenced by other factors, such as the duration or timing of fixations, which may provide more insights into the driver's attentional allocation and comprehension of the driving situation.

The negative correlation between the dispersion of fixations on the game and SA was consistent with the idea that excessive visual engagement with NDRT degraded SA (Du, Yang, & Zhou, 2020; Zhou et al., 2021). Greater dispersion on the game surface may indicate increased distraction and reduced focus on the driving environment, leading to lower SA. The positive correlation between dispersion of fixations on the center area and higher SA aligned with previous findings (Du, Yang, & Zhou, 2020; Liang, Yang, Yu, et al., 2021). A wider distribution of fixations on the road ahead and surrounding areas can facilitate better perception and comprehension of the driving situation, thereby enhancing SA.

Overall, the integration of machine learning techniques with multimodal datasets presents a powerful approach to uncover the nuanced factors that influence SA in the realm of automated driving. Through continual refinement of feature sets and comparison with extant research, we might not only be able to advance the predictability of SA levels, but also understand the underlying cognitive and behavioral processes involved.

7.2   Effects of risks and automation errors on SA

In this study, we manipulated participants' risk perceptions through exposure to risk-related content and varied the performance of the automated driving system and the simulated driving environment. The results revealed that when participants were exposed to high-risk content, their SA levels were higher. These findings are supported

by previous studies that demonstrated that ((M. Li et al., 2019; Yahoodik & Yamani, 2021)) perceived risks and trust in an AV were affected by introductory information and received training related to system reliability.

Notably, SA was significantly sensitive to automation performance and was higher when the system experienced failures compared to error-free drives. This observation is consistent with studies indicating that automation errors and system limitations can trigger compensatory behaviors and heightened alertness in drivers, leading to improved SA during critical situations (Körber et al., 2018; Schwarz et al., 2019)

The eye tracking responses supported this observation, as participants spent a significant amount of time gazing at the road center when driving in risk conditions (see Fig. 8 b) and experienced automation errors (see Fig. 8 c). Moreover, the high risk condition resulted in notably increased dispersion on road center (see Fig. 8 e) specifically in erroneous situations (see Fig. 8 f), as well as left and right sides (see Fig. 8 i and l). This broader visual scanning pattern suggests that individuals may explore the traffic environment more extensively and scan across a wider range of locations when faced with higher perceived risks or system failures, potentially in an attempt to gather more information and enhance their understanding of the situation (de Winter et al., 2019; Thill et al., 2014).

Furthermore, the data from the eye tracker revealed that participants checked the traffic around them more frequently in presence of automation error (see Fig. 8 g) and were less focused on the NDRT screen (see Fig. 8 h). This shift in attentional allocation from the NDRT to the road scene during automation errors is consistent with previous findings that drivers tend to prioritize monitoring the driving environment over secondary tasks when faced with critical situations or system limitations (Naujoks, Mai, & Neukum, 2014; Smith et al., 2023).

Finally, we observed that the participants demonstrated longer and more dispersed gazing on the odometer when they experienced failures (see Fig. 8 j and k), which was noted when participants noticed the hazard ahead before they received a response from the AV. This behavior might indicate that they were trying to understand if the

automation was going to react to the potential hazard or checking if the automation was turned off after they resumed control (Niu et al., 2022; Zeeb, Buchner, & Schrauf, 2015). Overall, these findings contribute to our understanding of how risk perception and automation reliability influence drivers' visual attention allocation, situation scanning strategies, and cognitive processes related to SA in conditionally automated driving scenarios. The results align with and extend the existing literature on the effects of perceived risk, system limitations, and critical events on drivers' vigilance, compensatory behaviors, and SA in human-automation interaction contexts.

7.3 Implications

We developed a context-aware model using machine learning to predict SA during conditionally AVs, which could have significant implications for enhancing safety and user experience. First, by combining diverse data sources like demographics, physiological signals, and eye movements, the model can capture a more comprehensive understanding of the driver's state and attentional focus. This holistic approach could lead to more accurate predictions of SA levels during automated driving scenarios, which could provide deeper insights into the driver's cognitive processes and readiness to take over control when needed (Du et al., 2019). Moreover, incorporating demographic factors like age, gender, and driving experience into the model allows for personalized driver monitoring and tailored interventions (Avetisyan, Ayoub, Yang, & Zhou, 2023). By accurately predicting SA levels, the model can inform the design of adaptive in-vehicle systems and human-machine interfaces (HMIs) that provide context-aware support and warnings to drivers (Pakdamanian et al., 2022). This could lead to safer and smoother transitions between automated and manual driving modes. The model's predictions can also guide the development of personalized training programs or adaptive automation strategies, helping drivers maintain an appropriate level of SA and readiness during automated driving (Du et al., 2019).

Overall, developing a context-aware ML model that leverages multimodal data sources has the potential to significantly enhance SA prediction and support safer and more

user-friendly conditionally automated driving experiences. However, technical challenges, data availability, and ethical considerations must be carefully addressed to realize the full benefits of this approach.

7.4   Limitations and Future Work

This study has several limitations that should be acknowledged. First, the experimental setup used a low-fidelity desktop driving simulator, which may not fully replicate the realistic dynamics and risk perceptions of an on-road driving environment. Future studies should aim to conduct experiments in higher-fidelity simulated or real-world settings to enhance the ecological validity of the findings.

Second, the takeover scenarios tested in this study covered a limited range of risk levels. To gain a more comprehensive understanding of SA level, it is essential to explore a broader spectrum of risky situations, including more critical and time-sensitive takeover events.

Third, the study relied on a single self-reported item to measure SA, which may not fully capture the multidimensional nature of this construct. While this choice was made to minimize intrusion during the driving task and prioritize the exploration of the LightGBM model's predictive capabilities, a separate validation study for the single-item questionnaire was not conducted which inherently linked the model's performance to the reliability of the self-reported SA measure used as ground truth. Future research could enhance the robustness of SA measurement by exploring alternative, more comprehensive self-report instruments, or by validating the single-item measure against established SA assessments and objective performance indicators, thereby strengthening the foundation for model development and evaluation.

Finally, integrating diverse data sources and developing robust ML models for SA prediction can be technically challenging, requiring advanced data fusion techniques and large, high-quality datasets for training and validation. The model's performance and reliability in real-world driving scenarios need to be thoroughly evaluated, as factors like environmental conditions and unexpected events can influence SA and driver behavior.

8   Conclusions

This paper presents research on developing a predictive model for assessing SA in conditionally automated driving scenarios. An experiment with 67 participants using a driving simulator was conducted. Participants experienced automated driving with TOR events, including some with automation errors (i.e., false alarms and misses). Their physiological responses (GSR, HR, eye tracking) and self-reported SA were recorded. The LightGBM machine learning model was used to predict SA levels from the physiological and demographic data. The model achieved reasonable performance ($RMSE = 0.89, MAE = 0.71, Corr = 0.78$) using a subset of the top 12 most important features resulted from SHAP explainer. The key findings were: 1) age, AV knowledge, GSR, HR, and eye behavior on the center and NDRT screens were the most influential predictors of SA, 2) higher risk perception led to larger fixations durations and dispersions on center screen, 3) automation errors increased the dispersions and fixations on center and NDRT screens, and 4) SA was higher during automation error conditions for the high risk group compared to low risk.

These findings contribute to our understanding of the factors influencing SA in conditionally automated driving scenarios, particularly considering the impact of risk perception and automation errors. The developed predictive model demonstrates the potential for using physiological, behavioral and demographic measures to monitor and assess drivers' SA in real-time, enabling intelligent vehicle systems to provide timely interventions or explanations to enhance SA and promote safer human-AV interactions.

9   Acknowledgement

10   References

Avetisyan, L., Ayoub, J., Yang, X. J., & Zhou, F. (2023). Building trust profiles in conditionally automated driving. *arXiv preprint arXiv:2306.16567*.

Ayoub, J., Avetisian, L., Yang, X. J., & Zhou, F. (2023). Real-time trust prediction in conditionally automated driving using physiological measures. *IEEE Transactions on Intelligent Transportation Systems*.

Ayoub, J., Yang, X. J., & Zhou, F. (2021). Modeling perceived trust in automated vehicles with predictability and explainability. *Transportation Research Part F: Psychology and Behaviour*.

Ayoub, J., Zhou, F., Bao, S., & Yang, X. J. (2019). From manual driving to automated driving: A review of 10 years of autoui. In *Proceedings of the 11th international conference on automotive user interfaces and interactive vehicular applications* (pp. 70–90).

de Winter, J. C., Eisma, Y. B., Cabrall, C., Hancock, P. A., & Stanton, N. A. (2019). Situation awareness based on eye movements in relation to the task environment. *Cognition, Technology & Work*, *21*(1), 99–111.

Du, N., Ayoub, J., Zhou, F., Pradhan, A. K., Robert, L., Tilbury, D., . . . Yang, X. J. (2019). Examining the impacts of drivers' emotions on takeover readiness and performance in highly automated driving. In *Proceedings of the human factors and ergonomics society annual meeting* (Vol. 63, pp. 2076–2077).

Du, N., Yang, X. J., & Zhou, F. (2020). Psychophysiological responses to takeover requests in conditionally automated driving. *Accident Analysis & Prevention*, *148*, 105804.

Du, N., Zhou, F., Pulver, E. M., Tilbury, D. M., Robert, L. P., Pradhan, A. K., & Yang, X. J. (2020a). Examining the effects of emotional valence and arousal on takeover performance in conditionally automated driving. *Transportation research part C: emerging technologies*, *112*, 78–87.

Du, N., Zhou, F., Pulver, E. M., Tilbury, D. M., Robert, L. P., Pradhan, A. K., & Yang, X. J. (2020b). Predicting driver takeover performance in conditionally

automated driving. *Accident Analysis & Prevention*, *148*, 105748.

Durso, F. T., Hackworth, C. A., Truitt, T. R., Crutchfield, J., Nikolic, D., & Manning, C. A. (1998). Situation awareness as a predictor of performance for en route air traffic controllers. *Air Traffic Control Quarterly*, *6*(1), 1–20.

Endsley, M. R. (1988). Design and evaluation for situation awareness enhancement. In *Proceedings of the human factors society annual meeting* (Vol. 32, pp. 97–101).

Endsley, M. R. (1995). Toward a theory of situation awareness in dynamic systems. *Human Factors*, *37*(1), 32-64.

Endsley, M. R., & Kiris, E. O. (1995). The out-of-the-loop performance problem and level of control in automation. *Human factors*, *37*(2), 381–394.

Fernandez Rojas, R., Debie, E., Fidock, J., Barlow, M., Kasmarik, K., Anavatti, S., . . . Abbass, H. (2019). Encephalographic assessment of situation awareness in teleoperation of human-swarm teaming. In *Neural information processing: 26th international conference, iconip 2019, sydney, nsw, australia, december 12–15, 2019, proceedings, part iv 26* (pp. 530–539).

Holthausen, B. E., Wintersberger, P., Walker, B. N., & Riener, A. (2020). Situational trust scale for automated driving (sts-ad): Development and initial validation. In *12th international conference on automotive user interfaces and interactive vehicular applications* (pp. 40–47).

Hulse, L. M., Xie, H., & Galea, E. R. (2018). Perceptions of autonomous vehicles: Relationships with road users, risk, gender and age. *Safety science*, *102*, 1–13.

Jensen, T., Khan, M. M. H., Albayram, Y., Fahim, M. A. A., Buck, R., & Coman, E. (2020). Anticipated emotions in initial trust evaluations of a drone system based on performance and process information. *International Journal of Human–Computer Interaction*, *36*(4), 316–325.

Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., . . . Liu, T.-Y. (2017). Lightgbm: A highly efficient gradient boosting decision tree. *Advances in neural information processing systems*, *30*.

Khastgir, S., Birrell, S., Dhadyalla, G., Sivencrona, H., & Jennings, P. (2017). Towards

increased reliability by objectification of hazard analysis and risk assessment (hara) of automated automotive systems. *Safety Science*, *99*, 166–177.

Kintz, J. R., Banerjee, N. T., Zhang, J. Y., Anderson, A. P., & Clark, T. K. (2023). Estimation of subjectively reported trust, mental workload, and situation awareness using unobtrusive measures. *Human Factors*, *65*(6), 1142–1160.

Körber, M., Baseler, E., & Bengler, K. (2018). Introduction matters: Manipulating trust in automation and reliance in automated driving. *Applied ergonomics*, *66*, 18–31.

Kuhn, H. W., & Tucker, A. W. (1953). *Contributions to the theory of games* (No. 28). Princeton University Press.

Li, M., Holthausen, B. E., Stuck, R. E., & Walker, B. N. (2019). No risk no trust: Investigating perceived risk in highly automated driving. In *Proceedings of the 11th international conference on automotive user interfaces and interactive vehicular applications* (pp. 177–185).

Li, S., Blythe, P., Guo, W., & Namdeo, A. (2018). Investigation of older driver's takeover performance in highly automated vehicles in adverse weather conditions. *IET Intelligent Transport Systems*, *12*(9), 1157–1165.

Li, Y., Zhao, L., Gao, K., An, Y., & Andric, J. (2022). Revealing driver psychophysiological response to emergency braking in distracted driving based on field experiments. *Journal of intelligent and connected vehicles*, *5*(3), 270–282.

Liang, N., Yang, J., Prakah-Asante, K. O., Curry, R., Blommer, M., Swaminathan, R., ... Pitts, B. J. (2021). Look up! an eye-tracking study on situation awareness during automated vehicle takeover. In *Proceedings of the human factors and ergonomics society annual meeting* (Vol. 65, pp. 1071–1071).

Liang, N., Yang, J., Yu, D., Prakah-Asante, K. O., Curry, R., Blommer, M., ... Pitts, B. J. (2021). Using eye-tracking to investigate the effects of pre-takeover visual engagement on situation awareness during automated driving. *Accident Analysis & Prevention*, *157*, 106143.

Lundberg, S. M., Erion, G., Chen, H., DeGrave, A., Prutkin, J. M., Nair, B., ... Lee,

S.-I. (2020). From local explanations to global understanding with explainable ai for trees. *Nature machine intelligence*, *2*(1), 56–67.

Lundberg, S. M., & Lee, S.-I. (2017a). A unified approach to interpreting model predictions. In I. Guyon et al. (Eds.), *Advances in neural information processing systems 30* (pp. 4765–4774). Curran Associates, Inc. Retrieved from `http://papers.nips.cc/paper/` `7062-a-unified-approach-to-interpreting-model-predictions.pdf`

Lundberg, S. M., & Lee, S.-I. (2017b). A unified approach to interpreting model predictions. *Advances in neural information processing systems*, *30*.

Makowski, D., Pham, T., Lau, Z. J., Brammer, J. C., Lespinasse, F., Pham, H., . . . Chen, S. H. A. (2021, feb). NeuroKit2: A python toolbox for neurophysiological signal processing. *Behavior Research Methods*, *53*(4), 1689–1696. Retrieved from `https://doi.org/10.3758%2Fs13428-020-01516-y` doi: 10.3758/s13428-020-01516-y

Manchon, J., Beaufort, R., Bueno, M., & Navarro, J. (2022). Why does the automation say one thing but does something else? effect of the feedback consistency and the timing of error on trust in automated driving. *Information*, *13*(10), 480.

Merat, N., & Jamson, A. H. (2009). Is drivers' situation awareness influenced by a fully automated driving scenario? In *Human factors, security and safety.*

Naujoks, F., Mai, C., & Neukum, A. (2014). The effect of urgency of take-over requests during highly automated driving under distraction conditions. *Advances in Human Aspects of Transportation*, *7*(Part I), 431.

Niu, J., Zhang, Z., Sun, Y., Wang, X., Ni, J., & Qin, H. (2022). The driver's instantaneous situation awareness when the alarm rings during the take-over of vehicle control in automated driving. *Traffic injury prevention*, *23*(8), 478–482.

Pakdamanian, E., Hu, E., Sheng, S., Kraus, S., Heo, S., & Feng, L. (2022). Enjoy the ride consciously with cawa: Context-aware advisory warnings for automated driving. In *Proceedings of the 14th international conference on automotive user interfaces and interactive vehicular applications* (pp. 75–85).

Perello-March, J. R., Burns, C. G., Birrell, S. A., Woodman, R., & Elliott, M. T. (2022). Physiological measures of risk perception in highly automated driving. *IEEE Transactions on Intelligent Transportation Systems*, *23*(5), 4811–4822.

Perello-March, J. R., Burns, C. G., Woodman, R., Elliott, M. T., & Birrell, S. A. (2021). Driver state monitoring: Manipulating reliability expectations in simulated automated driving scenarios. *IEEE transactions on intelligent transportation systems*, *23*(6), 5187–5197.

Pop, V. L., Shrewsbury, A., & Durso, F. T. (2015). Individual differences in the calibration of trust in automation. *Human factors*, *57*(4), 545–556.

SAE. (2021). *Taxonomy and definitions for terms related to driving automation systems for on-road motor vehicles*. SAE International in United States, *J*3016_202104.

Salmon, P., Stanton, N., Walker, G., & Green, D. (2006). Situation awareness measurement: A review of applicability for c4i environments. *Applied ergonomics*, *37*(2), 225–238.

Salvucci, D. D., & Goldberg, J. H. (2000). Identifying fixations and saccades in eye-tracking protocols. In *Proceedings of the 2000 symposium on eye tracking research & applications* (pp. 71–78).

Schwarz, C., Gaspar, J., & Brown, T. (2019). The effect of reliability on drivers' trust and behavior in conditional automation. *Cognition, Technology & Work*, *21*(1), 41–54.

Seppelt, B. D., & Lee, J. D. (2019). Keeping the driver in the loop: Dynamic feedback to support appropriate use of imperfect vehicle control automation. *International Journal of Human-Computer Studies*, *125*, 66-80.

Siennicka, A., Quintana, D., Fedurek, P., Wijata, A., Paleczny, B., Ponikowska, B., & Danel, D. (2019). Resting heart rate variability, attention and attention maintenance in young adults. *International Journal of Psychophysiology*, *143*, 126–131.

Smith, K., Endsley, T., & Clark, T. (2023). Physiological correlates of objective situation awareness measurements. In *2023 ieee aerospace conference* (pp. 1–6).

Taylor, R. M. (2017). Situational awareness rating technique (sart): The development of a tool for aircrew systems design. In *Situational awareness* (pp. 111–128). Routledge.

Thill, S., Hemeren, P. E., & Nilsson, M. (2014). The apparent intelligence of a system as a factor in situation awareness. In *2014 ieee international inter-disciplinary conference on cognitive methods in situation awareness and decision support (cogsima)* (pp. 52–58).

Unni, A., Ihme, K., Jipp, M., & Rieger, J. W. (2017). Assessing the driver's current level of working memory load with high density functional near-infrared spectroscopy: A realistic driving simulator study. *Frontiers in human neuroscience*, *11*, 167.

Yahoodik, S., & Yamani, Y. (2021). Effectiveness of risk awareness perception training in dynamic simulator scenarios involving salient distractors. *Transportation research part F: traffic psychology and behaviour*, *81*, 295–305.

Yeo, L. G., Sun, H., Liu, Y., Trapsilawati, F., Sourina, O., Chen, C.-H., ... Ang, W. T. (2017). Mobile eeg-based situation awareness recognition for air traffic controllers. In *2017 ieee international conference on systems, man, and cybernetics (smc)* (pp. 3030–3035).

Zeeb, K., Buchner, A., & Schrauf, M. (2015). What determines the take-over time? an integrated model approach of driver take-over after automated driving. *Accident analysis & prevention*, *78*, 212–221.

Zhang, T., Tao, D., Qu, X., Zhang, X., Lin, R., & Zhang, W. (2019). The roles of initial trust and perceived risk in public's acceptance of automated vehicles. *Transportation research part C: emerging technologies*, *98*, 207–220.

Zhang, T., Yang, J., Liang, N., Pitts, B. J., Prakah-Asante, K., Curry, R., ... Yu, D. (2023). Physiological measurements of situation awareness: a systematic review. *Human factors*, *65*(5), 737–758.

Zheng, B., Jin, X., Song, Z., Pei, Y., & Ma, X. (2018). Effect of different information push mechanism on driver's situation awareness. In *Engineering psychology and*

cognitive ergonomics: 15th international conference, epce 2018, held as part of hci international 2018, las vegas, nv, usa, july 15-20, 2018, proceedings 15 (pp. 250–262).

Zhou, F., Alsaid, A., Blommer, M., Curry, R., Swaminathan, R., Kochhar, D., . . . Tijerina, L. (2022). Predicting driver fatigue in monotonous automated driving with explanation using gpboost and shap. *International Journal of Human–Computer Interaction*, *38*(8), 719–729.
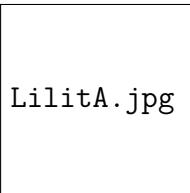
Zhou, F., Qu, X., Helander, M. G., & Jiao, J. R. (2011). Affect prediction from physiological measures via visual stimuli. *International Journal of Human-Computer Studies*, *69*(12), 801–819.

Zhou, F., Qu, X., Jiao, J., & Helander, M. G. (2014). Emotion prediction from physiological signals: A comparison study between visual and auditory elicitors. *Interacting with computers*, *26*(3), 285–302.

Zhou, F., Yang, X. J., & De Winter, J. C. (2021). Using eye-tracking data to predict situation awareness in real time during takeover transitions in conditionally automated driving. *IEEE Transactions on Intelligent Transportation Systems*, *23*(3), 2284–2295.
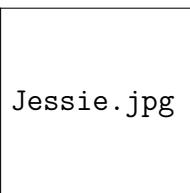
Zhou, F., Yang, X. J., & Zhang, X. (2019). Takeover Transition in Autonomous Vehicles: A YouTube Study. *International Journal of Human–Computer Interaction*, *0*(0), 1–12. doi: 10.1080/10447318.2019.1634317

11   Biographies

LilitA.jpg

Lilit Avetisyan received her B.E. degree in 2017 and MS degree in 2019 in Information Security from the National Polytechnic University of Armenia. She is currently pursuing her Ph.D. degree in Industrial and Systems Engineering at the University of Michigan, Dearborn. Her main research interests include human-computer interaction, explainable artificial intelligence and situation awareness.

Jessie.jpg

X. Jessie Yang is an Assistant Professor in the Department of Industrial and Operations Engineering, University of Michigan, Ann Arbor. She earned a PhD in Mechanical and Aerospace Engineering (Human Factors) from Nanyang Technological University, Singapore. Dr. Yang's research include human-autonomy interaction, human factors in high-risk industries and user experience design.

Feng.jpg

Feng Zhou received the Ph.D. degree in Human Factors Engineering from Nanyang Technological University, Singapore, in 2011 and Ph.D. degree in Mechanical Engineering from Gatech Tech in 2014. He was a Research Scientist at MediaScience, Austin TX, from 2015 to 2017. He is currently an Assistant Professor with the Department of Industrial and Manufacturing Systems Engineering, University of Michigan, Dearborn. His main research interests include human factors, human-computer interaction, engineering design, and human-centered design.