Building a Situation Awareness Framework for Communicating Among

Road Users in Mixed Automated and Manual Traffic Environments

by

Lilit Avetisyan

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Industrial and Systems Engineering)
in the University of Michigan-Dearborn
2024

Doctoral Committee:

Associate Professor Feng Zhou, Chair
Assistant Professor Areen Alsaid
Associate Professor Shan Bao
Associate Professor Xi Jessie Yang, University of Michigan

Lilit Avetisyan

lilita@umich.edu

ORCID iD: 0000-0003-4206-6385

Dedication

In loving memory of my dear uncles, Dr. Samvel and Manvel Avetisyan, whose wisdom and dedication to knowledge have been inspiration for me throughout my academic journey.

To my parents and siblings, whose unconditional love and steadfast belief in me have made this achievement possible.

To my cherished extended family, a constant source of encouragement and strength throughout this challenging endeavor.

To the mentors and friends who have challenged me intellectually and supported me personally.

Table of Contents

List of Figures

List of Tables

List of Abbreviations

ANOVA Analysis of Variance

AV Autonomous vehicle

CV Conventional vehicle

DALI Driving Activity Load Index

EEG Electroencephalography

eHMI external human-machine interface

GOSS Gradient-based One-Side Sampling

GSR Galvanic Skin Response

HMI Human-Machine Interface

HR Heart Rate

HRV Heart Rate Variability

iHMI internal human-machine interface

LightGBM Light Gradient Boosting Machine

MAE Mean Absolute Error

NDRT Non-driving related tasks

PPG Photoplethysmography

RMSE Root Mean Square Error

RMSSD Root Mean Square of Successive Differences

SA Situation Awareness

SAE Society of Automotive Engineers

SAGAT Situation Awareness Global Assessment Technique

SART Situational Awareness Rating Technique

SAT SA-based Agent Transparency

SD Standard Deviation

SHAP SHapley Additive exPlanations

SKR Skill-Knowledge-Rule

SPAM Situation Present Assessment Method

SSQ Simulator Sickness Questionnaire

STS-AD Situational Trust Scale for Automated Driving

TLX Task Load Index

TOR Take-Over Request

VR Virtual Reality

VRU Vulnerable road user

V2X Vehicle-to-Everything

XAI Explainable artificial intelligence

Abstract

On the promise of reducing human error and increasing road safety, the number of autonomous vehicles (AVs) in the industry and on the roads is steadily rising. However, this technology faces challenges in various scenarios, such as interactions with AV drivers or in mixed traffic environments where AVs share the road with conventional vehicles (CV), pedestrians, and bicyclists. In these situations, situation awareness (i.e., the ability to perceive, comprehend, and predict the situation on the road) is crucial to ensure road safety for all road users. To address this challenge, this dissertation aims to design and evaluate explanations based on the theory of mind to improve human-machine performance. The central hypothesis is that explanations during human-machine interaction can provide the necessary information to resume situation awareness, improving the joint performance of the human-machine team.

To achieve this objective, three fundamental questions were investigated: (1) how should the AV assess the driver's situation awareness to understand the need for explanations, (2) how should the AV driver understand the state of mind of the AV through explanations, and (3) how should the AV share information with CV drivers through explanations. To tackle these questions, first, a machine learning model was developed to predict the situation awareness of AV drivers in real-time using behavioral, physiological, and self-reported data. A LightGBM (Light Gradient Boosting Machine) model trained on the most critical predictors identified by SHAP (SHapley Additive exPlanations) achieved promising performance. Next, an explanation framework was proposed based on the situation awareness model and explainable AI. The framework was tested in an online environment by evaluating participants' sit-

uational trust, cognitive workload, and explanation satisfaction. The results showed that properly designed explanations based on the proposed framework assisted drivers in unexpected situations, increased their trust in AVs, and improved their situation awareness. Finally, external and internal HMI concepts were proposed to explore the interaction between AVs and CVs in challenging situations and improve situation awareness among CV drivers. The concepts were tested in a virtual reality environment using self-reported and physiological measurements. The findings revealed that explanations were able to increase participants' situation awareness and trust in AVs, with the internal HMI perceived as the most effective.

Overall, this research aims to contribute new fundamental knowledge about how to build situation awareness to improve human-machine performance by designing and evaluating human-centered explanations, particularly in conditionally automated driving. The research also seeks to enhance the relationship between humans and technology in automated driving and other fields, such as manufacturing and medical industries.

Chapter 1 Introduction

This chapter establishes the foundation of the research and introduces the research problem. Through reviewing the background information, this research is identified as a situation awareness framework for communication between road users in mixed automated and manual traffic environments, intended to propose a novel communication form that enhances the efficiency, interaction, and safety of autonomous and conventional vehicles (CVs). Additionally, this chapter outlines the research objectives, scope, and technical structure of the thesis.

1.1 Problem Statement

As autonomous vehicles (AVs) become more common on the roads, a mixed traffic, where AVs share the road with other road users, is expected to become more prevalent than the conventional traffic environment dominated by human-driven vehicles. This transition of environment poses a number of challenges that need to be addressed in order to support the mixed traffic, increase public understanding of AV technology and ensure safe and efficient cooperation between road users. One of the main challenges that AVs face in mixed traffic is communication (*Bhavsar et al.,* 2017), encompassing both intra-vehicle communication to facilitate cooperation between the AV and its human operator, and inter-vehicle communication to foster safe and effective interaction with other road users, including CVs and vulnerable road users (VRU), such as pedestrians and bicyclists. While increasing the automation

level tries to enhance transportation safety, vehicles alone do not ensure complete safety. Overall, safety of AVs is heavily reliant on the human-AV team performance. Hence, it is crucial to enhance the communication in order to provide road users the necessary information to make quick, informed decisions, enabling them to react to any potentially dangerous situations as they arise. The decreased level of situation awareness (SA) among road users in mixed traffic is a primary factor that necessitates communication. This decrease negatively impacts their ability to perceive, comprehend, and predict the situation on the road, leading to a range of critical issues such as difficulties in establishing trust among road users, acceptance and use of AV technologies, and the potential for serious and even fatal accidents (*Choi and Ji*, 2015). Nevertheless, it is important to emphasize that road users (i.e., AV operator, CV driver, VRUs, infrastructure) in the mixed traffic environment differs with their communication method, and have their specific requirements and needs.

1.1.1 Intra-vehicle Communication

Numerous studies have been carried out to understand how the level of automation impacts AV drivers, and emphasized that the advancement of automation level in vehicles (*Society of Automotive Engineers*, 2018) shifts the responsibility of AV drivers from control to monitoring or assisting, leading to a reduction in attention resources allocated to driving and an increase in focus on non-driving tasks (*De Winter et al.*, 2014; *Endsley*, 2018; *Frison et al.*, 2019b). This occurrence results in a notable decline in AV drivers' SA, which pushes them out of the control loop and impairs their ability to assess the AV's actions accurately. Although the AV's promise to provide safe transportation, effective communication between the AV and its driver is crucial for exchanging information, including informing the driver about its intentions, interpreting the perceived information about their surrounding and alerting of any potential hazards, and maintaining "in-the-loop" SA level to safely and efficiently

operate the vehicle. By perceiving and comprehending the vehicle's surroundings, the AV driver can make informed decisions and take appropriate actions when interventions are needed (*Clark et al.*, 2017).

Several studies have attempted to address this problem by providing explanations of AV behavior and highlighting the drivers' need for information to be back in the control loop with the AV (*Koo et al.*, 2015, 2016; *Petersen et al.*, 2019; *Du et al.*, 2021). However, there is currently no direct evidence to support the effectiveness of such information in improving intra-vehicle communication, SA, and ultimately human-AV team performance. Furthermore, there is currently a lack of standardized models that can comprehensively evaluate the impact of proposed communication methods while introducing a minimal level of cognitive workload.

1.1.2 Inter-vehicle Communication

During the transition period when AVs are not yet widespread, the interaction between AVs and VRUs becomes more critical. AVs should be designed to coexist with other road users and accommodate their limitations. Unlike CVs, AVs are designed to analyze the behavior of CVs and predict their intentions to facilitate smoother interactions. Using advanced sensors and algorithms, AVs can anticipate the movements of road users and react accordingly in unexpected situations. However, in ambiguous or dynamic intersections (i.e., four-way, intersections, T-junctions without traffic lights and equal narrow passages) where the vehicles should negotiate the right-of-way, CV drivers are facing difficulties to understand the AVs intentions which negatively affects the traffic flow and human-drivers' trust in AV. In such situations, the AVs need to communicate with CV drivers via explicit signage and standardized communication protocols to share the awareness and assist CV drivers understand how to interact with AVs on the road. On the other hand, AV needs to safely collaborate with multiple road users, including CV drivers and VRUs, which can lead

to conflicts between road users due to the emergence of complex behavioral processes.

Previous studies have shown that communications through eHMIs (*Papakostopoulos et al.*, 2021; *Rettenmaier et al.*, 2020, 2019; *Eisma et al.*, 2021), as well as auditory signals and projections (*Bai et al.*, 2021; *Palmeiro et al.*, 2018; *Rasouli and Tsotsos*, 2019; *Colley and Rukzio*, 2020) could lead to better human-AI team performance. However, employment of these methods has been limited due to issues such as poor visibility at a distance, AV driver's engagement in NDRT and identifiability among other road users. In contrast to AV-VRU communication, the current understanding of AV-CV interactions is limited by the scarcity of investigations. The complex nature of these interactions necessitates a deeper exploration of various factors, including communication protocols, behavior prediction models, safety considerations, and the implications of mixed traffic scenarios. As a result, communication between AVs and CVs remains underexplored.

In human-AV interactions, improving AV transparency and transferring its knowledge through communication can potentially facilitate shared awareness. However, each road user employs a unidirectional communication method, leading to difficulties in understanding each other's mental models. Hence, I propose to build a situation awareness framework for communicating among road users in mixed automated and manual traffic environments, i.e., a communication method/structure for AVs applicable to both AV and CV drivers so the AV can share the situation awareness in a less cognitively demanding format for humans.

1.2 Research Objective

The primary objective of this dissertation is to build a communication framework for AVs and other road users of the human-AV team that will 1) establish SA dynamically in real time between the AV and road users and 2) improve the SA of both AV drivers and CV drivers in mixed traffic.

In pursuit of this objective, the following research questions are formulated:

1) How to establish situation awareness? To narrow down this research question in the context of mixed traffic, we broke it down into three sub-questions.

    1) What are the common informational needs of different road users in a human-AV in mixed traffic?

    2) How should this information be presented to different road users in mixed traffic?

    3) When should this information be presented to different road users in mixed traffic?

2) How to improve SA for AV drivers in mixed traffic?

3) How to improve SA for CV drivers in mixed traffic?

4) How to assess dynamic situation awareness in real-time?

Accordingly, the following research tasks are derived from the research questions:

1) Investigate the theoretical foundations of SA, theory-of-mind networks and explainable artificial intelligence (XAI), and build a framework for SA;

2) Examine the effect of communication with SA framework on AV drivers in mixed traffic using self-reported measurements;

3) Examine the effect of communication with SA framework on CV drivers in mixed traffic using self-reported and eye-tracking measurements;

4) Develop a computational model to predict drivers' situational awareness in real-time using physiological measurements.

1.3 Dissertation Outline

This dissertation comprises seven chapters, as shown in Fig. 1.1 that investigate the effects of situation awareness on communication between AVs and other road users, including their human drivers and CV drivers in mixed traffic scenarios.

Chapter 1 provides an overview of the research by discussing the general context, significance of the research topic, and summarizing the objectives of the dissertation.

Chapter 2 presents a comprehensive review of prior research concerning the communication challenges between AVs and road users. The chapter examines the primary factors causing issues in their coexistence, in particular SA, as well as the existing methods of improving SA in mixed traffic.

Chapter 3 lays the groundwork for constructing the SA framework by addressing the core challenges of human-AV communication. The chapter outlines a proposed methodology to overcome these issues and facilitate cooperation between AV and other road users in mixed traffic.

Chapter 4 presents the development of a real-time SA measurement model based on physiological metrics such as eye-tracking, GSR and HR. The chapter outlines an experiment that was conducted in desktop driving simulator to predict SA in real-time and to identify the most significant factors that contribute to inferring SA.

Chapter 5 investigates communication between AV and its drivers using explanations in conditional and highly automated driving using the developed situation awareness framework. The study evaluates the effectiveness of the framework across different levels of SA and information modalities, taking into account metrics such as trust, cognitive workload, and explanations satisfaction.

Chapter 6 reports the investigation of communication between AV and CV drivers using explanations based on the situation awareness framework in mixed traffic scenarios in a VR environment. In this research, the effectiveness of explanations is evaluated across different HMI conditions based on SA, trust, cognitive workload,

and acceptance metrics.

Finally, Chapter 7 provides a summary of the research that has been accomplished towards the stated research objectives, identifies areas for future work necessary to complete the remaining research goals, and highlights the limitations of the research that can be addressed in the future studies.
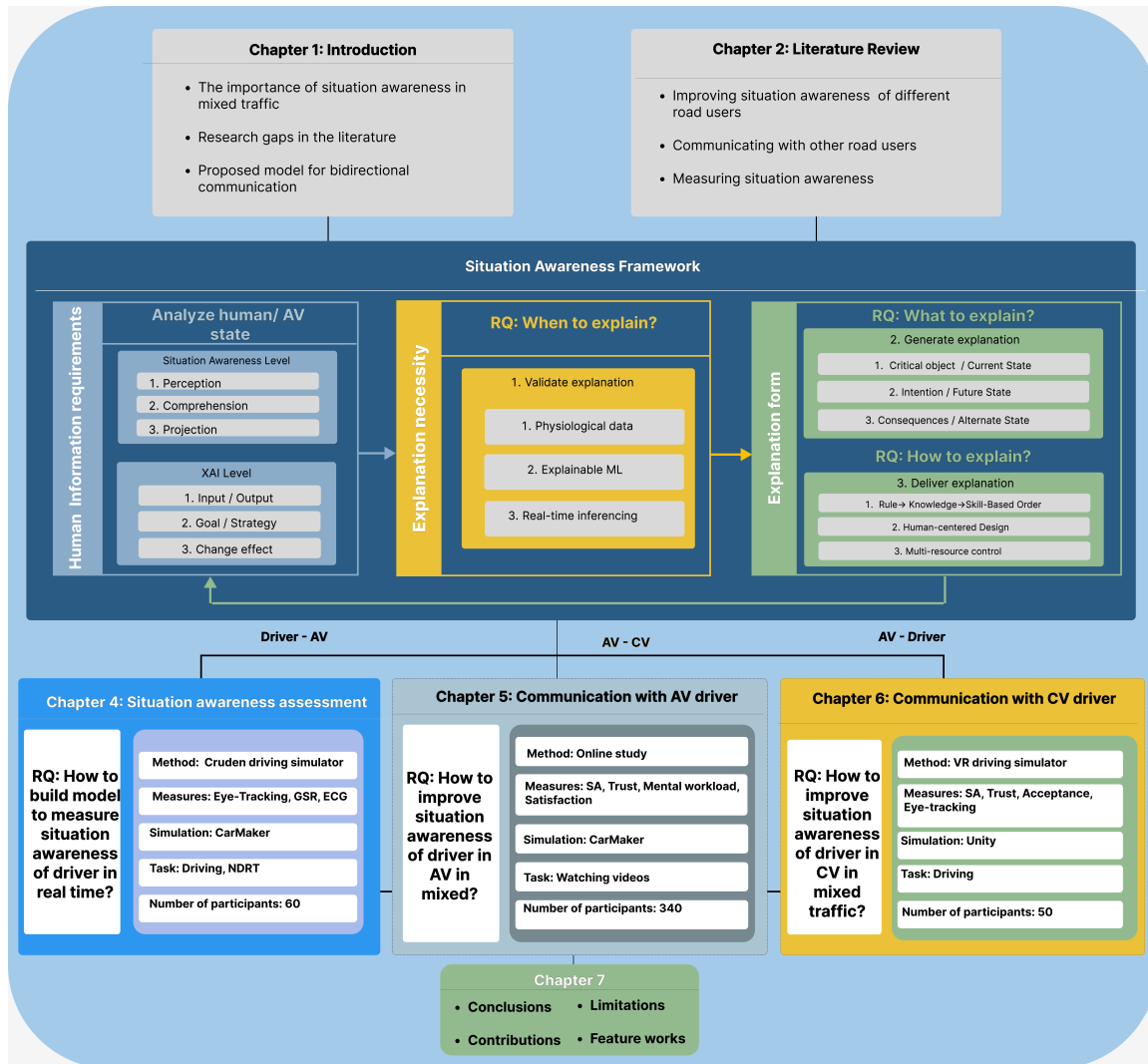
Figure 1.1: Research scope.

Chapter 2 Literature Review

2.1 Situation awareness and the out-of-the-loop problem

As the level of automation in vehicles increases, such as in conditional and highly AVs, i.e., SAE Levels 3 and 4 AVs (*SAE*, 2021), drivers' responsibilities shift from active operators to passive passengers for the majority of the time. Consequently, drivers become increasingly disengaged from the control loop, which reduces their SA since their attention is primarily diverted to NDRTs, resulting in less time spent monitoring the road and compromising their performance when intervention is required (*Frison et al.*, 2019a; *Endsley*, 2019). Although driver intervention may not be necessary in most situations, improving drivers' SA in unexpected driving scenarios is crucial for enhancing their trust in and acceptance of AVs. *Merat et al.* (2019) differentiated three kinds of loops in AV systems and described them as follows: 1) A driver was in the control loop when he/she was both in the physical control and monitoring the driving task, 2) a driver was on the control loop when the driver was only monitoring the driving task, and 3) a driver was out of the control loop as long as he/she was not monitoring the driving task. Thus, the out-of-the-loop problem in AVs describes the situation when the driver is not actively monitoring the system or the environment (*Radlmayr et al.*, 2014). This issue is mostly due to driver's overtrust in AVs, since a certain level of "control" is needed to properly respond to situational changes or to reduce uncertainty in automated driving, such as monitoring

and takeover control (*Du et al.*, 2020a, 2019a, 2020b).

Merat et al. (2019) emphasized that a key aspect to be in the control loop was the drivers' attention and cognitive responses to the changes in the system and in the dynamic environment, which was characterized by the driver's SA. In other words, when the driver is not in the control loop of the AV, the SA of system status and the driving environment may be reduced (*Sebok and Wickens*, 2017; *Zhou et al.*, 2019a, 2021b). Even if the driver is on the control loop (i.e., not in physical control of the vehicle, but monitoring the driving situation) (*Merat et al.*, 2019), he/she becomes a passive information processor, which would negatively affect the operator's understanding and comprehension (SA Level 2) of dynamic changes in the system even though the driver is aware of low-level information (SA Level 1) (*Endsley and Kiris*, 1995). This is further aggravated by the black-box decision-making process of the AV and the monotonicity of automated driving, which lead to low vigilance and even drowsiness (*Zhou et al.*, 2020, 2021a). However, SAE Levels 3-4 AVs allow drivers to conduct NDRTs without monitoring the driving task (*Ayoub et al.*, 2019a). In order to resolve such conflicts (i.e., conducting NDRTs in AVs vs. requiring a certain level of SA in AVs), explanations are needed to help drivers resume their SA in time when a certain level of "control" or understanding is needed to respond the situational changes, especially during unexpected driving scenarios.

2.1.1 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 as-

sess their level of awareness (*Endsley*, 1995a). 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 the limitations of subjective measures, researchers have explored the use of objective and continuous measures based on physiological and eye-tracking data. One such measure is Electroencephalography (EEG), which measures brainwave activity to provide insights into cognitive workload, attentional focus, and mental fatigue – all of which are related to situation awareness (SA) (*Fernandez Rojas et al.*, 2019; *Yeo et al.*, 2017). Another measure is Heart Rate Variability (HRV), where variations in heart rate reflect changes in the autonomic nervous system, influenced by cognitive and emotional states relevant to SA (*Perello-March et al.*, 2021). The Galvanic Skin Response (GSR) measure is also used, as it reflects changes in skin conductance associated with arousal and emotional responses that can indicate changes in SA (*Smith et al.*, 2023). Additionally, by monitoring eye movements, researchers can analyze eye-tracking metrics such as fixation duration, saccade patterns, and scan paths to understand where drivers focus their visual attention and how efficiently they process 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.

2.1.2 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. One notable approach involves deep learning models, which can effectively predict driver SA with high accuracy by analyzing physiological and behavioral data. For instance, *Li et al.* (2022) proposed a deep learning model that achieved promising results in predicting driver SA during conditionally automated driving scenarios. Another approach utilizes eye-tracking based models, as eye movements provide valuable insights into drivers' visual attention and information processing. *Zhou et al.* (2021b) developed a machine learning model using eye-tracking data to predict SA levels in real-time. Neuroimaging techniques, such as functional near-infrared spectroscopy (fNIRS), allow 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. Additionally, context-aware models incorporate information about the driving environment, traffic conditions, and driver behavior to provide more accurate and personalized predictions of SA. *Zheng 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.

2.2 Explanations for AV driver: Keeping the driver in the loop and maintaining situational awareness

In human factors research, explanations about the AV's behavior, system feedback and status, and driving scenarios were designed and provided to improve the transparency of system decisions and driver trust. For instance, Wintersberger et al. (2019) showed that augmented reality by coding traffic objects and future vehicle actions increased automation transparency and improved user trust and acceptance. *Koo et al.* (2015) designed three different types of information to explain AV behavior about: 1) "how" the car was acting, 2) "why" the car was acting and 3) "how" + "why" the car was acting. Authors investigated AV-driver interaction in a scenario where the AV took control from the driver and suddenly braked to avoid collision with an obstacle. They explained the AV behavior before the AV started acting, and found that "how" + "why" information resulted in the safest AV-driver cooperation , but also produced the greatest cognitive workload than other explanations, which could lead to confusion and anxiety. The "how" only information led to worse driving performance and unsafe cooperation since the drivers tried to take the control back from the AV but did not understand why the AV behaved in that way. Mackay et al.'s (2019) investigation into different amounts of feedback found that "more information does not necessarily lead to more trust and may, in fact, negatively affect cognitive load". Taehyun et al. (2020) stated that type of explanation significantly affects trust in AVs and suggested an explanation format based on the attribution theory (*Weiner*, 1979). They found that perceived risk moderated the effect of explanations on trust, i.e., attributional explanations led to the highest level of trust in low perceived risk compared to no or simple explanations.

In addition, the timing of the explanations (i.e., before or after particular action) also plays an important role in trust and acceptance in AVs. For example, Körber et al. (2018b) provided explanations of the causes of takeover requests after the takeover

transitions, which led to no decrease in trust or acceptance, but improved participants' understanding of system behaviors. *Koo et al.* (2015) argued that explanations should be provided ahead of an event which also was supported by Haspiel et al. (2018) and Du et. al. (2019b) studies, who found that explanations provided before the AV's action promoted more trust than those provided afterward. Thus, it is recommended that we should provide explanations before the vehicle takes action.

Other types of factors, such as forms, contents, and modalities of the explanations also play important roles in explanations in AVs. Wang et al. (2020) explored how information modality influenced driver's performance and showed that both visual and auditory modalities had a significant influence, but on different aspects of driver's performance. In particular, visual information boosted performance efficiency and auditory information decreased reaction time. *Seppelt and Lee* (2019) showed that continuous feedback helped drivers to be involved in the loop of system performance and operations. Consistent with the multiple resource theory (*Wickens*, 2008b), they found that the combined visual-auditory interface performed the best regarding drivers' confidence and trust.

2.3 Explanations for other road users: Communicating with human drivers and pedestrians

The integration of AVs into existing roadways poses numerous challenges (*Brown et al.*, 2023), particularly when they need to share the road with CVs and vulnerable road users, such as pedestrians and bicyclists. This coexistence goes beyond simply maneuvering within a physical environment; it necessitates that AVs interpret and accurately respond to the complex social dynamics of human behaviors on the road. Moreover, the growing prevalence of AVs necessitates exploring existing communication conventions and modifying them to address these challenges.Vinkhuyzen and Cefkin (2016) took an ethnographic approach to understand how AVs could be

developed to be socially acceptable and highlighted that human drivers use a range of subtle, non-verbal cues to negotiate traffic situations, which AVs would need to replicate to be perceived as "social actors" on the road. This aligns with Risto et al.'s (2017) emphasis on vehicle movement gestures and underscores the need for AVs to not just follow traffic rules, but to engage in socially understood behaviors. Youssef et al. (2024) explored factors affecting driver decision-making in narrow passage interactions and confirmed that besides traffic rules, drivers rely on a complex interplay of factors, including visibility, vehicle size, and perceived intentions of other drivers, to make yielding decisions. A recent study delved deeper into the cognitive processes behind drivers' overtaking decisions when encountering oncoming AVs. It found that these decisions are shaped by factors such as perceived time-to-collision, the AV's behavior, and provides insights into how AVs could communicate their intentions to influence CV drivers' decisions *Mohammad et al.* (2024).

In certain situations, such as four-way intersections (*Papakostopoulos et al.*, 2021) and pedestrian crosswalks (*Fuest et al.*, 2018), highway merging, bottleneck scenarios, effective communication and cooperation between AVs and other road users becomes necessary to ensure safe and efficient traffic flow, and communication methods that have clear in purely human-driven contexts may prove inadequate for this new paradigm of mixed traffic. Miller et al. (2022) conducted a mixed traffic driving simulator study to investigate how drivers' expectations about AVs and CVs impact their behavior and found that drivers who expected AVs to be more cautious and rule-abiding tended to behave more aggressively towards them, while those who expected CVs to be more unpredictable drove more defensively around them. Zhang et al. (2018) evaluated concepts for AV external communication systems, and suggests that while explicit communication has its place, it should be designed to mimic the simplicity and intuitiveness of the non-verbal cues that human drivers use. Several studies addressed these issues by improving technical aspects of the AV system.

Siebinga et al. (2023) propose that by integrating human communication and negotiation strategies into AV systems, these vehicles can better predict and respond to human drivers' actions, thereby reducing the risk of accidents and conflicts. Isele et al. (2018) addressed a key challenge in AV navigation where the AV's sensors cannot fully perceive the environment and developed a deep reinforcement learning approach that allowed AVs to navigate such intersections safely by incorporating the intention of other drivers. However, Zgonnikov et al. (2024) argued that besides making intelligent decisions in challenging situations, AVs should be able to negotiate their right of way, and provided insights into the effectiveness of implicit communication by AVs. They demonstrated that AVs could "push" human drivers into desired behaviors by adjusting their own speed and position, without explicit signals. In contrast, other studies emphasized that cultural variability also poses a significant challenge for the design of AV communication systems, especially for vehicles that will operate in diverse international contexts (*Zhang et al.*, 2018; *Ehrhardt et al.*, 2024). Since such communication is subject to varying interpretations across different cultures, it becomes more complex to predict and generalize how they will be received and understood, making it challenging to establish a standard for such indirect forms of interaction. Aoki and Rajkumar (2018) investigated dynamic intersection scenarios with possible collision risks and showed that using V2V communications and sensor systems, AVs can effectively communicate with other road users and positively impact traffic flow. Rettenmaier et al. (2020) examined communication strategies between AVs and CVs in bottleneck situations where passage is restricted to a single vehicle. Their findings suggest that direct visual displays outperform laser projections, particularly in terms of visibility at greater distances (*Eisma et al.*, 2021). In another investigation, using directional arrows significantly enhanced comprehensibility, transferability, and simplicity, resulting in shorter passing times and fewer crashes compared to other eHMI designs (*Rettenmaier et al.*, 2019). Papakostopou-

15

los et al. (2021) investigated how eHMIs affect drivers' ability to infer AV motion intentions and found that eHMIs accelerated CV drivers' decision-making, improved driving behavior, and reduced overall crossing times. Although limited studies have explored HMIs in AV-CV communication, numerous studies demonstrate that eHMIs enhance interactions between AVs and VRUs. However, the specific content of these displays remains a subject for further investigation. Insights from Merat et al. (2018) suggested a preference for signals that communicate intended vehicle actions (e.g., initiating movement, turning, stopping) rather than data regarding vehicle dynamics (e.g., speed). Schieben et al. (2019) emphasized the importance of designing AV's intentions based on human needs and expectations, highlighting the need for AVs to adapt to other road users' behaviors. The design of eHMIs should also consider factors such as trust, acceptance, and user experience. Eisma et al. (2021) found that eHMIs using the receiver's perspective (egocentric) were more effective in communicating vehicle intent than those using an allocentric perspective. Conversely, Lee et al. (2021) found that certain eHMI designs could lead to confusion and mistrust, resulting in longer decision times and more errors. These findings underscore the importance of carefully designing HMIs to ensure they effectively communicate AV intentions without negatively impacting other road users' behavior, and confirm that HMIs may have a positive impact on CV driver's behavior as well.

Chapter 3 Building a Situation Awareness Framework

3.1 Introduction

Maintaining an appropriate level of SA is crucial for all road users in a mixed traffic environment, not only about their surroundings but also about other road users' knowledge and understanding. In human-AV teams, accurately predicting others' intentions allows for faster and safer decision-making and effective cooperation that can be achieved by building a team mental model (*Cannon-Bowers et al.*, 1993). All road users in such teams, i.e., AV, AV driver, CV driver and VRUs, share a common goal of safe navigation. However, each type of road users has their own sub-goals that require specific information to make decisions on how to proceed.

In the complex AI systems, such as AV, the decision-making process requires collecting and analyzing large amounts of data, and in a dynamic environment where AV needs to interact with various road users, such as mixed traffic, it is a challenge to decide which information to share.

Level 1: which pertains to "what" the AI system did or is doing

Level 2: which concerns "why" the system behaves in a particular manner or the significance of Level 1 information for the system

Level 3: which relates to "what" the system will do or what would happen "if" some of the inputs changed

Chen et al. (2014) proposed SA-based Agent Transparency (SAT) framework for dynamic environments that defines transparency requirements for AI systems by mapping SA levels as follows:

Level 1: system's status, purpose and intentions to explain "what" is going on and what is the trying to achieve

Level 2: reasons to explain "why" system is behaving in that way

Level 3: outcomes to explain "what" is the system's predictions for the future state

Previous studies have primarily focused on the sub-goals and used these frameworks to provide unidirectional explanations tailored to each user's mental model. For instance, in intra-vehicle communication, researchers aimed to efficiently facilitate takeovers, and provided "how", "why" and "what if" information about AV behavior to make the system transparent for AV drivers and boost the SA to necessary level. Similarly, in VRU-AV communication, researchers aimed to achieve efficient and fast street crossings, and provided mainly "how" information in various perspectives and formats, adjusted to the AV's or VRU's mental models.

In real-world scenarios, these varied explanations increase the complexity of AV behavior and require more communication with each road user to ensure optimal human-AV team performance. For instance, consider a scenario where an AV and a CV are approaching an intersection and the AV detects a pedestrian with an intention to cross the street. To coordinate the road users safely, the AV should yield to the pedestrian while simultaneously communicating with: 1) its driver to explain the reason for that behavior and prevent driver's interaction, 2) CV driver to share awareness about pedestrian and inform its intention, and 3) pedestrian to act as planned. This is critical because without this information, the AV driver's mental model may lead them to takeover and continue driving, the CV driver's mental model may lead them to cross the intersection, and the pedestrian's mental model may lead

18
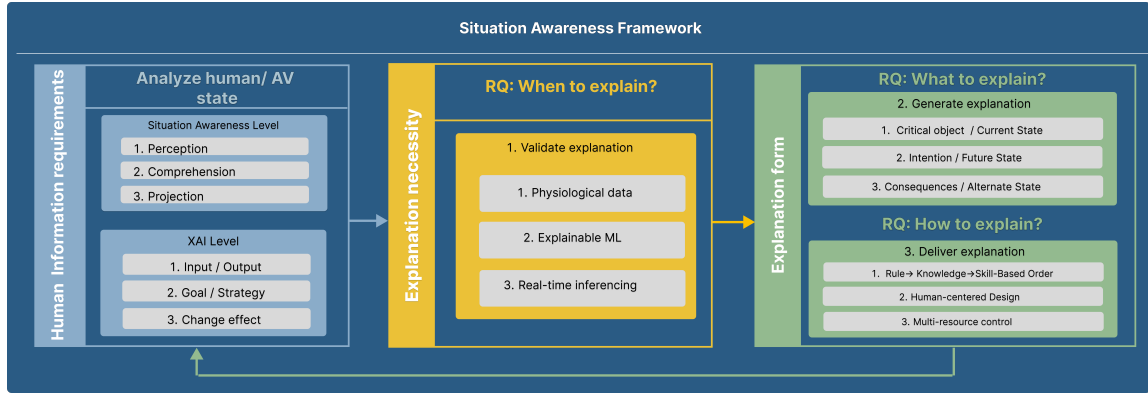
Figure 3.1: Situation awareness framework.

them to hesitate and negatively impact traffic flow, or cross blindly resulting in a potentially dangerous situation. However, if the AV understands the other road users' mental model, it can adjust their behavior accordingly. To achieve proactive communication and better team performance, it is essential to establish a shared mental model among road users. This necessitates the development of a situational awareness framework for AVs that can communicate in both directions (i.e, with AV drivers and with other road users) using the same format.

3.2 Situation Awareness Framework

Drawing upon previous models and study results, we propose a framework for building SA among human-AV teams in mixed traffic environments, as shown in Fig. 3.1. We defined the SA as the interface for AVs to share their awareness with other road users regarding the current, next, and alternate states of the environment. To ensure the effectiveness of the framework, we identified three key questions that must be addressed:

1) What are the common informational needs of different road users in a human-AV in mixed traffic?

2) How should this information be presented to different road users in mixed traf-

19

fic?

3) When should this information be presented to different road users in mixed traffic?

3.2.1 Generating explanations

To address the first question and fulfill the cognitive requirements of the SA model (*Endsley*, 1995a), we propose a SA framework with three levels as follows:

1. Level 1: Critical element in the mixed traffic environment - This level involves identifying a single critical object in the environment that has the highest importance in triggering the decision-making process for all human-AV team members. According to XAI framework principles (*Sanneman and Shah*, 2020), this object can be the system classification output. However, we suggest using the current state of the object instead of information about "what" the AV decided to do. For instance, in the intersection scenario mentioned earlier, the pedestrian can be the critical object, and the classified action can be "moving." It is vital for the pedestrians to know that they are being detected, and for the AV and CV to know of the pedestrian's presence. Overall, level 1 can be referred to as "moving pedestrian" for the current state of the environment in this example.

2. Level 2: Intention of road users - While level 1 conveys what is detected, level 2 conveys the classified actions to other road users for coordinating the situation. In the intersection scenario, for example, the AV can communicate with the pedestrian about its intention to yield by signaling them to cross, and also communicate with the CV driver by sharing the information that the pedestrian is crossing. Level 2 can be described as "crossing pedestrian" for the next state of the environment.

3. Level 3: Consequences - At this level, road users need to know the predicted

state of the environment if road users do not follow the shared plan. As the alternate step of the environment, level 3 can be described by "high risk of accident".

With these three levels, the proposed framework establishes the explanatory content that an AV can employ for effective communication with different categories of road users, (i.e., AV drivers, VRUs, and CV drivers). Additionally, the framework broadens the role of AVs within the traffic ecosystem, elevating them from mere road users into coordinators in challenging situations. However, it is crucial to deliver the content in an appropriate format to avoid confusion and prevent an increase in mental workload.

3.2.2 Delivering explanations

In human-machine interaction, selecting a communication interface that effectively conveys explanation content in a comprehensible manner presents a significant challenge. Even if the content itself is sufficient to support SA, its presentation can negatively impact overall performance by imposing additional cognitive load on recipients and causing trust issues. *Vicente and Rasmussen* (1992) emphasized the importance of designing the explanatory context according to the principles of ecological interface design (EID). They highlighted that information visualizations based on individuals' cognitive processing levels, as defined by the SKR taxonomy (*Rasmussen*, 1983), i.e., knowledge-based, skill-based, and rule-based processing, can enhance transparency and lead to faster, more accurate responses, while appropriately calibrating trust in the system. Knowledge-based visualizations are based on symbolic representations (i.e., text language), necessitating analytical problem-solving resources and which may vary depending on differences in individuals' mental models. On the other hand, rule-based visualizations (i.e., as sign language) and skill-based visualizations (i.e., sensory signal language) pertain to perception and action resources, enabling the

conveyance of information without necessarily relying on individuals' mental models. This allows for parallel and effortless processing of information, facilitating rapid and accurate reactions.

While all these processing levels are integral to the decision-making process, their prioritization hinges upon the mission requirements they serve. Previous research already showed the effectiveness of using the EID approach in AVs (*Stanton et al.*, 2021; *Chen et al.*, 2019; *Schewe and Vollrath*, 2020) in cognitive load and cooperative performance. However, in challenging human-AV interactions, communication with road users can be time-critical and may demand immediate responses from both ends. As this framework proposes multidirectional communication, it is essential to communicate with road users also in a mutually exclusive way. Therefore, to address the second question, we propose to display the explanations employing the SKR levels following a specific priority order:

1. Rule-based explanation: This level places emphasis on explaining the current, next, and alternate traffic states using universally recognized traffic signs. By leveraging these signs and associated rules with them, reliable cue-action behavior can be coordinated, enabling road users to navigate in the environment safely and efficiently with minimal cognitive load. For example, at Level 1, the interface should present critical components such as crossing pedestrians, malfunctioning traffic lights, and road accidents to describe the current situation in the environment. At Level 2, the interface should include signs that enforce specific behaviors, and as the AV communicates in different directions, the actions should always be presented from the perspective of the road user. For instance, displaying signs like "stop", "yield", "go", and "speed limit" can explain the expected behavior for CVs, while using symbols like a "walking person" or an "upraised hand" can convey the expected behavior for VRUs as the next state of the traffic environment. Similarly, Level 3 explanations can

22

utilize signs related to road accidents or traffic jams to depict alternate states of the situation.

2. Knowledge-based explanation: In order to support the comprehension and handling of unfamiliar rules, we suggest integrating a symbolic model of explanations, such as natural language texts, within the explanation interface. These textual explanations serve the purpose of assisting individuals in navigating novel situations and facilitating the development of a shared mental model among road users.

3. Skill-based explanation: To support faster and accurate information processing, we suggest providing explanations in different modalities such as visual (i.e, text or augmented reality) and audio (i.e, speech). Besides, allowing different modalities we can accommodate individuals with different sensory abilities. For people with visual impairments, audio information becomes crucial in understanding and navigating the environment. Similarly, individuals with hearing impairments can rely on visual explanations for comprehension. By communicating through multiple modalities, we ensure accessibility and inclusivity for a wider range of individuals.

3.2.3 Estimating necessity

Apart from determining which explanations to provide, it is equally important to understand when to provide them. In mixed traffic, situations can change rapidly and explanations that may be necessary at one moment may not be needed in another one depending on SA of each road users and situation complexity. Providing explanations to road users when they have adequate SA could potentially be bothersome and elicit an adverse response, as drivers may disregard the explanations which can have negative implications for the acceptance and safety of AVs. In other words, AVs

should claim the driver's attention only when it is necessary to enforce sufficient levels of SA for safe driving. Consequently, it becomes crucial to consistently evaluate drivers' level of SA and adapt the strategies used for explanations accordingly.

On the other hand, by continuously tracking the driver's state, we can validate the impact of explanations and make system adjustments to achieve desired outcomes. This approach also provides an opportunity to develop customized explanations in the future, leading to improved overall performance. It is important to note that cognitive skills vary among individuals, influenced by factors such as genetics, environment, education, and experiences. As a result, the level of detail required in explanations may differ, ultimately affecting decision-making time. For instance, experienced drivers may achieve the desired SA level with only Level 1 explanation, while providing them with all three levels would unnecessarily increase their cognitive workload. In contrast, inexperienced drivers may need to receive the full set of explanations to attain a sufficient SA level. Therefore, we propose to include a component in the explanation system that can objectively assess driver's SA in real time. This assessment will help determine the threshold at which an explanation becomes necessary and will contribute to system improvement by identifying the most important factors that characterize the driver's SA.

3.3 Summary

Recognizing the limitations of existing XAI and models in AV that primarily focus on unidirectional communication, we propose a SA framework that introduces a communication model for AV and enables effective intra- and inter-vehicle communication. Within this framework, we suggest a three-level structure for generating explanations that align with the informational needs for sufficient SA, along with an information delivery format for an effective cognitive control, and with a real-time SA assessment module for explanation timing and depth. The subsequent chapters

present applied examples of proposed SA framework in AV-driver and AV-CV driver interaction contexts.

Chapter 4 Towards Context-Aware Modeling of Situation Awareness in Conditionally Automated Driving

4.1 Introduction

The rapid advancement of AV technologies holds the promise of transforming transportation. Yet, as vehicles progress through the levels of automation set by the 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) (*Zhou et al.*, 2019b; *Ayoub et al.*, 2019b; *Zhou et al.*, 2021b). Research consistently demonstrates that delayed and ineffective driver response during takeovers, especially when distracted by NDRTs, substantially harm driving safety (*Du et al.*, 2019a, 2020b,a).

In conditionally automated vehicles, the maintenance of driver's SA, i.e. the accurate perception, comprehension, and projection of environmental elements (*Endsley*, 1995a), is paramount. The criticality of SA in ensuring timely and effective driver response during takeovers is well documented (*Salmon et al.*, 2006). However, driver engagement in NDRTs can lead to degradation of SA, leading to failures in takeovers (*Körber et al.*, 2018a; *Endsley and Kiris*, 1995).

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 et al.*, 2018; *Seppelt and 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 (*Pop et al.*, 2015; *Khastgir et al.*, 2017).

Given the transient dynamics of SA in the driving context, traditional SA measurement methods, while robust, prove inadequate for capturing the continuous evolution of SA, due to their intrusive nature and reliance on self-reports, such as freeze-probe techniques (e.g., SAGAT) or observer rating scales (e.g., SART) (*de Winter et al.*, 2019; *Durso et al.*, 1998). This knowledge gap highlights the need for developing unobtrusive, real-time methodologies that utilize objective indicators, such as physiological signals and eye-tracking data to assess SA.

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 (*Zhou et al.*, 2021b; *Du et al.*, 2020c). 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 et al.*, 2021; *Smith et al.*, 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 reliability conditions. Physiological data, i.e, GSR, HR, HRV 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 dis-

play 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 and 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 study are 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.

4.2 Method

4.2.1 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 were used for further analysis. All the participants had a valid driver's license with an average of 9.1 years of experience. Participants received

Figure 4.1: Experiment setup.

$30 in compensation for about 75 min of participation. The study was approved by the Institutional Review Board at the University of Michigan.

4.2.2 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 non-driving related tasks (NDRTs), and one phone, positioned to the participant's right side, for recording SA assessments (see Fig. 4.1). 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*, 2014). To engage in the automated 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. 4.3) developed on the Qualtrics platform (Provo, UT), 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, the time delays for each piece of equipment, including the iMotions, Pupil Core, driving simulator, and Tetris game, were recorded at the moment of the experiment's initiation.

4.2.3   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 as-

signed 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 (high-risk video: `https://www.youtube.com/watch?v=RC9iK1lV77E&t=4s`), coupled with an environment simulation of driving in foggy weather; 2) a low-risk condition, which shared positive feedback on AV performance via videos demonstrating AV's ability to anticipate safety-critical hazards that may be difficult for humans to detect (low-risk video: `https://www.youtube.com/watch?v=O4IUcOxXZqo`), and featured sunny weather driving simulation. 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. 4.4). Previous research suggested both false alarms and misses can decrease user trust (*Ayoub et al.*, 2023). The sequence of drives was counterbalanced across participants.

4.2.4 Experimental procedure

Fig. 4.2 provides an overview of the experiment 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 on 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. It was further explained that there might be instances where the AV could fail to detect an obstacle, termed as a "miss" condition, and situations where it might issue unnecessary TORs, referred to as "false alarms". Following the briefing, participants underwent

31

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. Once these devices were correctly configured, participants completed an online survey to provide demographic information. Subsequently, they were presented with information specific to their assigned risk condition. Next, participants were guided through a training session to familiarize themselves with both the driving simulator's functionality and the experimental protocol.
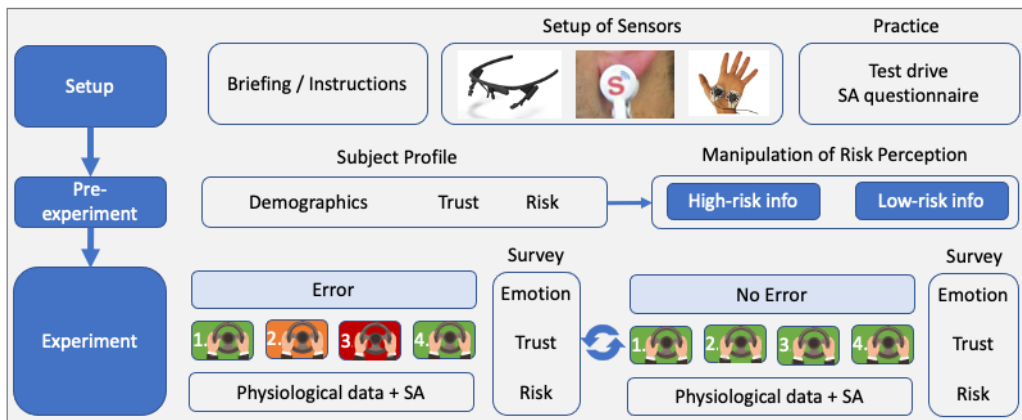
Figure 4.2: Experiment layout.

Finally, they proceeded to the actual driving sessions, each lasting approximately 15 minutes. During the drive, participants were asked to self-assess and report their SA levels every 30 seconds on a 4-item scale ranging from 0 to 3, with the instruction "Please indicate your situation awareness" (see Fig. 4.3). After each driving session, participants were requested to complete surveys evaluating their trust (*Holthausen et al.*, 2020b), emotional responses (*Jensen et al.*, 2020), and perceived risk (*Zhang et al.*, 2019) based on their recent driving experience using a 7-point Likert scale. 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 4.3: SA level assessment prompt.
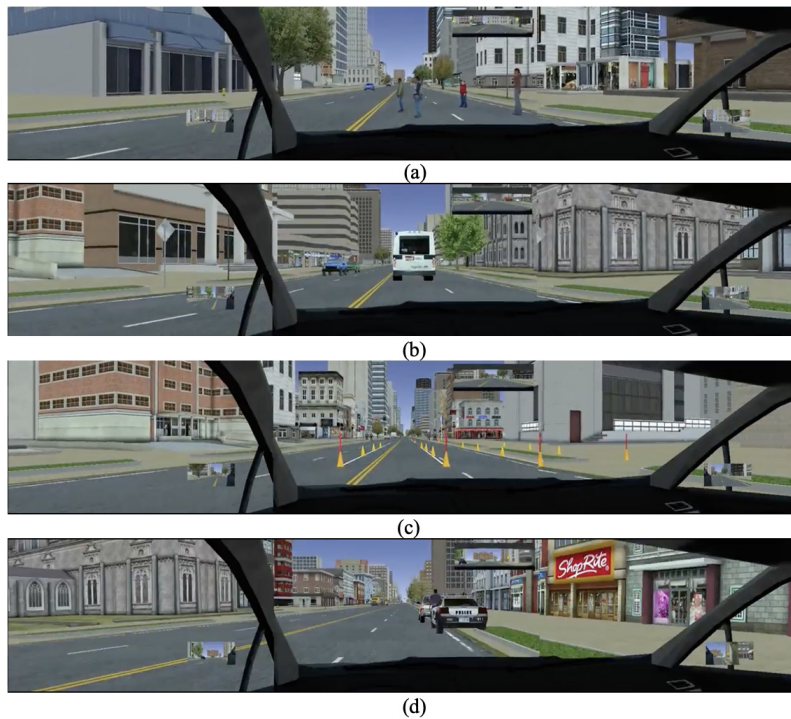
Figure 4.4: Takeover events in urban areas (a) pedestrians crossing ahead (b) bus sudden stop ahead (c) construction zone ahead (d) police vehicle on shoulder.

4.3 Predictive SA Model

4.3.1 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. Each component of data was processed as follows:

GSR Processing: The GSR signal was decomposed into tonic and phasic compo-

nents 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) with Root Mean Square of Successive Differences (RMSSD) method (see Eq. 1):

$$RMSSD = \sqrt{\sum_{i=1}^{N} \left(RR\_interval_i - RR\_interval_{i+1}\right)^2},$$ (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 and 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 and demographic data were linked to the physiological and eye-tracking data based on timestamps as shown in Fig. 4.5.

Integration: The final dataset integrated physiological, eye-tracking, and demographic data (age, gender, AV knowledge level), resulting in the 21 features outlined in Table 4.1.

4.3.2 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,
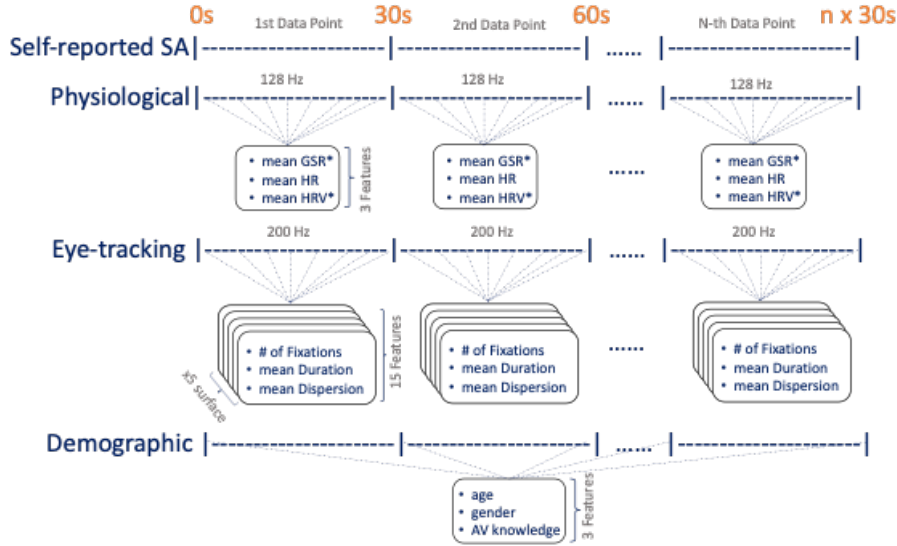
Figure 4.5: Data pre-processing and synchronization.

alongside self-reported assessments and demographic variables. After testing several ML algorithms, the LightGBM framework proved to be the most effective and was selected for this purpose.

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 the memory usage and improve the 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. Based on the tuning results, the following values 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. As critical indicators of model's performance, the root mean square error (RMSE) and mean absolute error (MAE) were calculated as follows:

$$RMSE = \sqrt{\frac{\sum_{i=1}^{N} \left(y_i - \hat{y}_i\right)^2}{N}}, \tag{2}$$

$$MAE = \frac{\sum_{i=1}^{N} |y_i - \hat{y}_i|}{N}, \tag{3}$$

$$Corr. = \frac{\sum_{i=1}^{N} (\hat{y}_i - \bar{\hat{y}})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{N} (\hat{y}_i - \bar{\hat{y}})^2 \sum_{i=1}^{N} (y_i - \bar{y})^2}}, \tag{4}$$

where N is the total number of the samples in the dataset, y_i is the i-th value of SA samples, $\hat{y}_i$ is the i-th predicted SA, $\bar{y}$ is the mean value of all the SA samples, and $\bar{\hat{y}}$ is the mean value of all the predicted SA results.

4.3.3 SHAP Explainer

To reveal the contribution of each feature in the predictive model of SA, the SHapley Additive exPlanations (SHAP) (*Lundberg and 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 and 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

Table 4.1: 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

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, 2021). The model training procedure is shown in Fig. 4.6
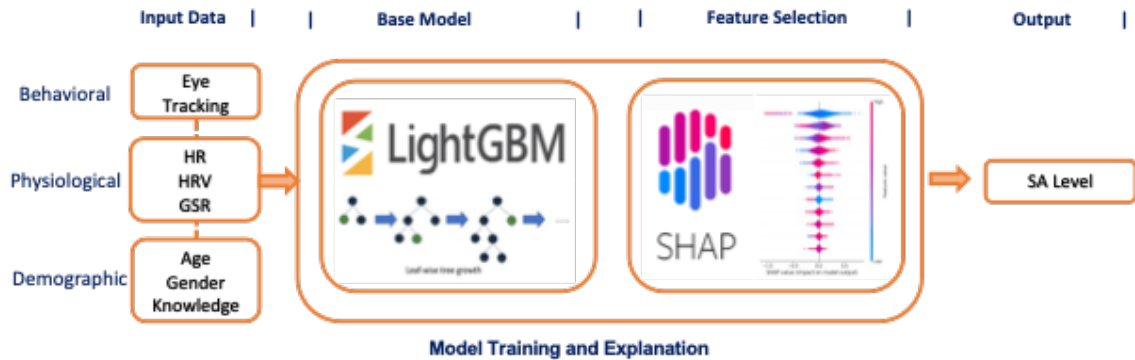
Figure 4.6: Steps for training the SA prediction model.

4.4 Results

4.4.1 Model performance

The LightGBM model's performance in predicting SA is presented in Table 4.2. To analyze how individual features contribute to predictions, SHAP values were calculated and visualized in a SHAP summary plot (Fig. 4.7). 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 very 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 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 importance rankings. The impact of each addition on the model's accuracy is illustrated in Fig. 4.8, which shows the the variations in performance metrics (i.e., RMSE, MAE and Corr) with the
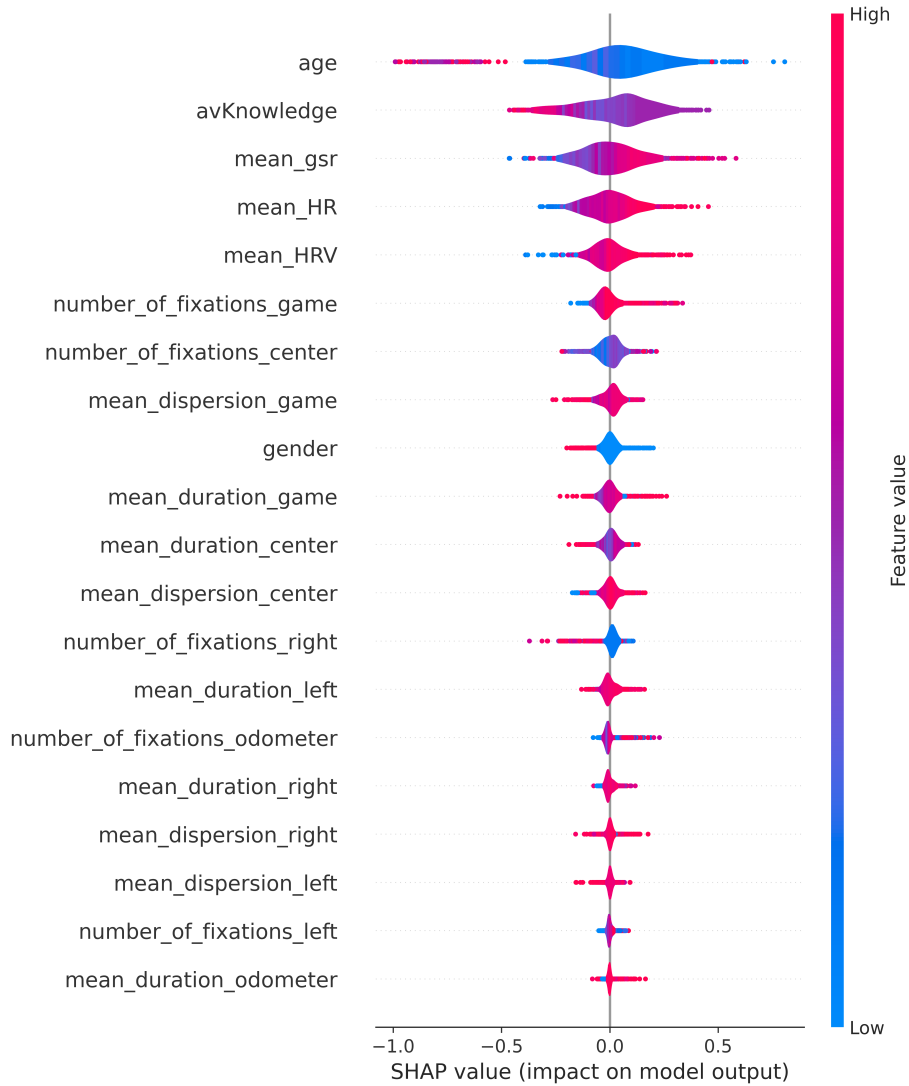
Figure 4.7: SHAP summary plot. The x-axis shows the feature's influence on SA. The y-axis shows the importance ranking of the features.

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's had the best performance (see Table 4.2 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, mean_dispersion_center. However, the summary plot shows only the global view of how feature values influenced predictions.
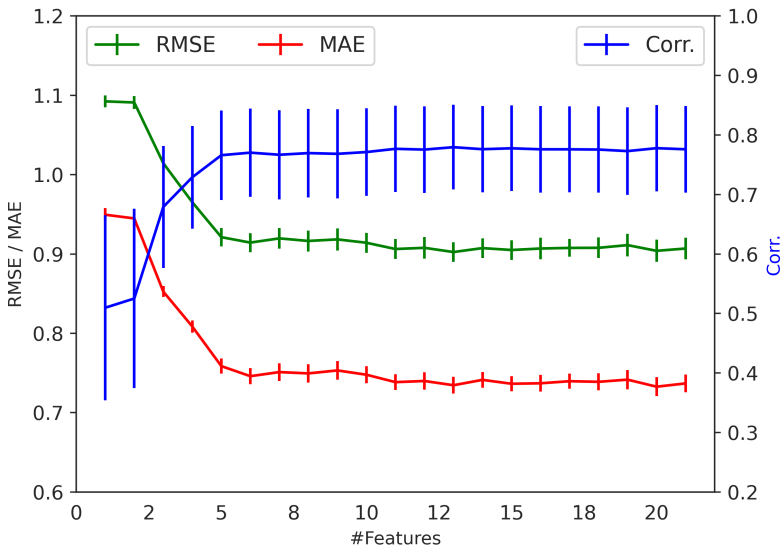
Figure 4.8: LightGBM model performance over the iteration of adding important featured at the time.

Table 4.2: Performance of LightGBM regressor.

Sample size	RMSE	MAE	Corr
All features	0.90	0.72	0.77
Selected Features	0.89	0.71	0.78

4.4.2 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 feature's actual values. This relationship is shown in Fig. 4.9, where the feature values were segmented into several groups according to their distribution (the physiological features were limited up 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.
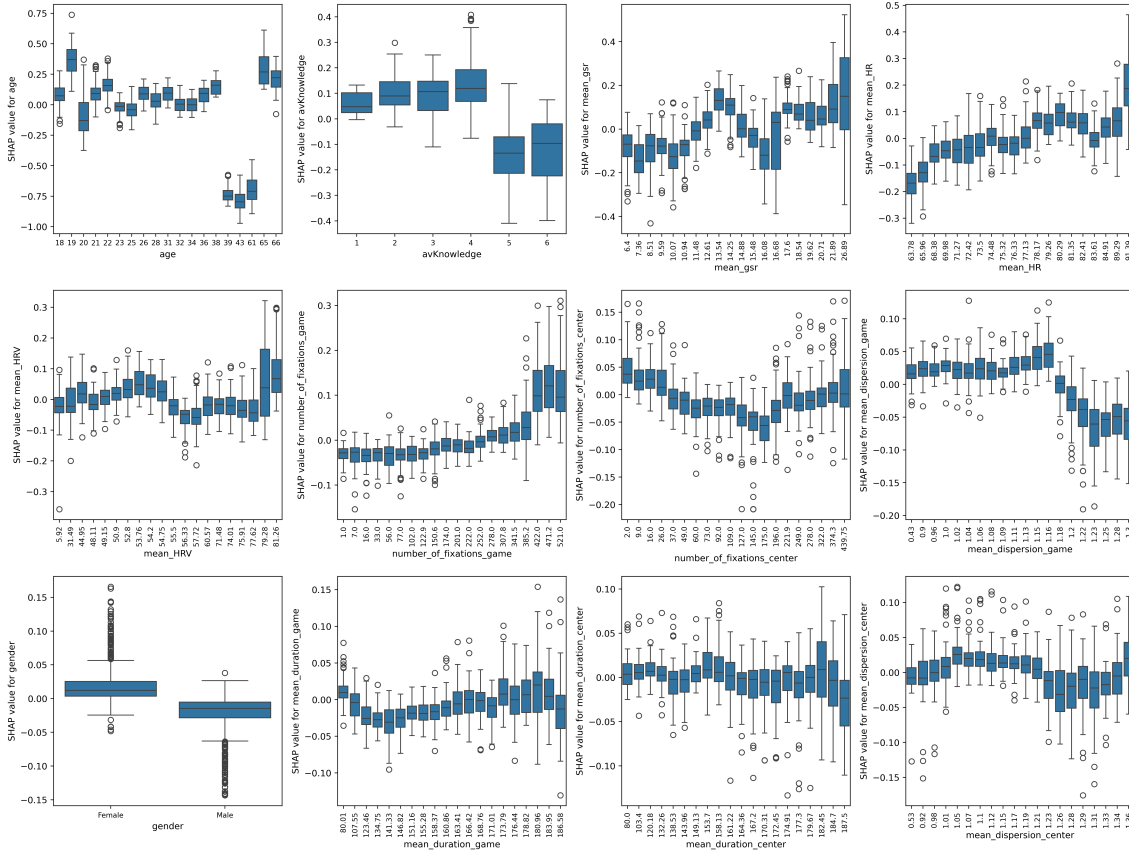
Figure 4.9: The effect of important features on predicted SA value. The x-axis represents the value of the feature, and the y-axis represents the SHAP value associated with that feature. Positive SHAP values indicate that the feature pushes the prediction higher, while negative values indicate the opposite.

Age: The most important feature was age in Fig. 4.7 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 significant 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 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 ($\rho = -0.670, p < 0.001$).

4.4.3 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 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 low-risk group. In terms of trust, a significant main effect of automation error was found ($F(1, 62) = 13.700, p < 0.001, \eta_p^2 = 0.181$). For risk perception, although no significant differences were found between the two conditions, it successfully elicit different levels of self-reported SA.

4.4.4 Objective Responses Across Conditions

The effect of risk and automation error was also investigated for physiological features using a two-way mixed ANOVA (see Table 4.3).

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 ($F = 5.34, p = 0.03, \eta_p^2 = 0.188$) and automation error ($F = 9.077, p = 0.006, \eta_p^2 = 0.283$), as well as a significant interaction effect ($F = 5.089, p = 0.034, \eta_p^2 = 0.181$). 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 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 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.
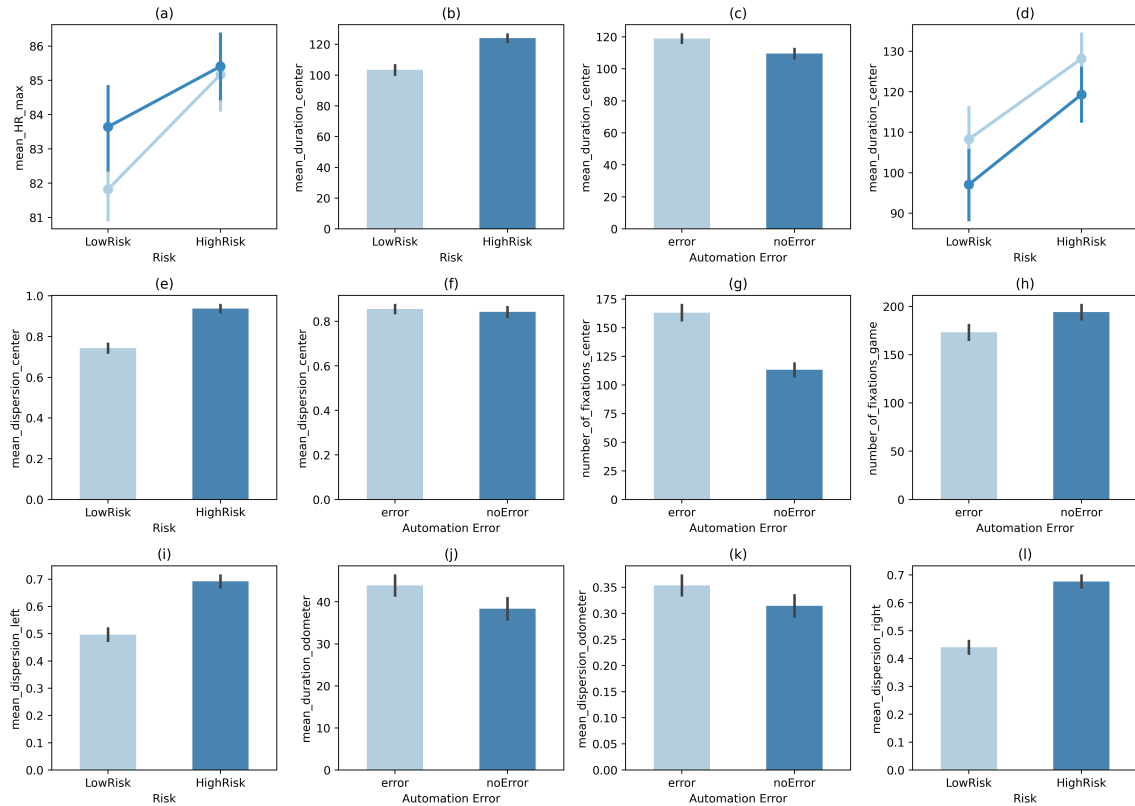
Figure 4.10: Main effects on features across tested conditions.

Table 4.3: Statistical results of physiological measures

Feature	Effect	F	p-value	Effect size
	Risk	0.106	0.748	0.005
mean_gsr	Automation Error	0.157	0.696	0.007
	Interaction	0.002	0.969	0.000
	Risk	3.47	0.075	0.131
mean_HR	Automation Error	0.034	0.986	0.000
	Interaction *	7.348	0.012	0.242
	Risk	2.245	0.148	0.089
mean_HRV				

44

	Automation Error	0.04	0.843	0.002
	Interaction	1.089	0.307	0.045
	Risk *	5.34	0.03	0.188
mean_duration_center	Automation Error *	9.077	0.006	0.283
	Interaction *	5.089	0.034	0.181
	Risk *	5.988	0.022	0.207
mean_dispersion_center	Automation Error *	4.657	0.042	0.168
	Interaction	3.653	0.069	0.137
	Risk	2.724	0.112	0.106
#_of_fixations_center	Automation Error *	10.786	0.003	0.319
	Interaction	0.011	0.918	0.000
	Risk	0.036	0.851	0.002
mean_duration_game	Automation Error	0.116	0.736	0.005
	Interaction	0.844	0.368	0.035
	Risk	0.509	0.483	0.022
mean_dispersion_game	Automation Error	0.297	0.591	0.013
	Interaction	0.562	0.461	0.024
	Risk	0.64	0.432	0.027
#_of_fixations_game	Automation Error *	8.423	0.008	0.268
	Interaction	0.118	0.734	0.005
	Risk	3.806	0.063	0.142
mean_duration_left	Automation Error	0.144	0.708	0.006
	Interaction	0.002	0.969	0.000
	Risk *	5.516	0.028	0.193
mean_dispersion_left	Automation Error	1.52	0.230	0.062
	Interaction	0.081	0.779	0.004
	Risk	3.346	0.08	0.127
#_of_fixations_left	Automation Error	2.794	0.108	0.108
	Interaction	0.591	0.450	0.025
	Risk	1.039	0.319	0.043
mean_duration_odometer	Automation Error *	4.698	0.041	0.170
	Interaction	0.189	0.668	0.008
	Risk	0.633	0.434	0.027
mean_dispersion_odometer	Automation Error *	6.817	0.016	0.229

	Interaction	0.002	0.964	0.000
	Risk	0.526	0.476	0.022
#_of_fixations_odometer	Automation Error	2.385	0.136	0.094
	Interaction	0.254	0.619	0.011
	Risk	3.872	0.061	0.144
mean_duration_right	Automation Error	1.821	0.190	0.073
	Interaction	1.319	0.263	0.054
	Risk *	4.41	0.047	0.161
mean_dispersion_right	Automation Error	0.898	0.353	0.038
	Interaction	0.858	0.364	0.036
	Risk	2.844	0.105	0.11
#_of_fixations_right	Automation Error	0.35	0.56	0.015
	Interaction	0.155	0.697	0.007

4.5   Discussion

4.5.1   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 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 (*Ay-*

oub 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 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 and confidence levels in tasks, thereby impacting SA in an automated context (*Kintz et al.*, 2023; *Li et al.*, 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 et al.*, 2011, 2014), which further lead 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 AVs, which was integral for the participants to pay more attention to the driving situations (*Perello-March et al.*, 2022). Previous studies also found that a higher value of HRV was associated with better attentional maintenance and flexibility, which might 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 (*Zhou et al.*, 2021b; *Liang et al.*, 2021a). 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 (*Zhou et al.*, 2021b; *Du et al.*, 2020a). 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 et al.*, 2020a; *Liang et al.*, 2021b). 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.

4.5.2   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 who demonstrated that (*Li et al.*, 2019; *Yahoodik and 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.*, 2018a; *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. 4.10 b) and experienced automation errors (see Fig. 4.10 c). Moreover, the high risk condition resulted in notably increased dispersion on road center (see Fig. 4.10 e) specifically in erroneous situations (see Fig. 4.10 f), as well as left and right sides (see Fig. 4.10 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 (*Thill et al.*, 2014; *de Winter et al.*, 2019).

Furthermore, the data from the eye tracker revealed that participants checked the traffic around them more frequently in presence of automation error (see Fig. 4.10 g) and were less focused on the NDRT screen (see Fig. 4.10 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 (*Smith et al.*, 2023; *Naujoks et al.*, 2014).

Finally, we observed that the participants demonstrated longer and more dispersed gazing on the odometer when they experienced failures (see Fig. 4.10 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 (*Zeeb et al.*, 2015; *Niu et al.*, 2022).

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.

4.5.3   Implications

We developed a context-aware model using machine learning to predict SA during conditionally AVs 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.*, 2019a). Moreover, incorporating demographic factors like age, gender, and driving experience into the model allows for personalized driver monitoring and tailored interventions (*Avetisyan et al.*, 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.*, 2019a).

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.

4.5.4   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. Future research should explore alternative methods for assessing situation awareness, such as objective performance measures or more comprehensive self-report instruments, to establish a more reliable ground truth.

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.

4.6 Conclusions

This study 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.

Chapter 5 The Impacts of Situation Awareness and Modality on Explanations in the Context of Conditional and Highly Automated Driving

5.1 Introduction

When drivers are out of the control loop, they will have a low level of SA, making it difficult for them to comprehend AV's behavior in unexpected situations. Moreover, it limits their ability to successfully take over control in critical situations, leading to accidents. For example, by analyzing Uber's AV fatal accident in Arizona (*Garcia*, 2018), it was revealed that the driver failed to take over control of the AV because she was engaged on her phone and was not aware of the pedestrian crossing the road. Regardless of who was responsible for the accident, such cases overall had negative impacts on trust in and public acceptance of AV. In particular, being unaware of the situation, drivers tend to interpret the AV's unexpected behavior as system malfunction that leads to trust issues in AVs. Hence, when the automated mode is on, the AVs should provide sufficient information to increase drivers' SA level for proper understanding of the situation and to ensure that the situation is under control.

Previous studies have addressed the SA and its associated issues in AVs (i.e., trust and takeover performance) through explanations, and provided important implications for designing AV systems. However, these solutions did not systematically assess how they improve drivers' trust with a minimal level of cognitive workload.

Therefore, it is necessary to theoretically frame explanations to support human-AV interaction.

In this study, we hypothesize that explaining AV behavior using our SA framework will fulfill drivers' informational needs and promote different levels of situation understanding with an optimal cognitive workload, resulting in improved human-AV performance. We expected that our explanation framework would foster trust with a relatively less increase in mental workload compared to the previous approaches due to the mapping of explanations to information processing levels. In order to test the hypothesis, we designed a three by two between-subjects experiment, where three types of explanations were manipulated to three levels of SA with two modalities (visual, visual + auditory) across six scenarios. We examined the effects of explanations in the form of three levels of SA on drivers' situational trust, cognitive workload, and explanation satisfaction.

5.2 Method

5.2.1 Participants

In total, 340 participants (151 females and 189 males; Age = 39.0 ± 11.4 years old) in the United States participated in this study. All the participants were recruited from Amazon Mechanical Turk (MTurk) with a valid US driver's license. On average, participants had 15 ± 11.8 years of driving experience and the driving frequency was 5 ± 1 days per week. They were randomly assigned to one of the seven conditions as shown in Table 5.1, where L1, L2, and L3 conditions were mapped closely to three SA levels proposed by Endsley. More detailed information about the experiment conditions is described in the "Scenario Design" section. This study was approved by the Institutional Review Board at the University of Michigan. Each participant was compensated with $2 upon completion of the study. The average completion time of the survey was about 26 minutes across the conditions.

Table 5.1: Experimental design with Modality and SA level as independent variables. The modality factor had two levels: 1) Visual, i.e., the explanation was given only in text format, and 2) Visual + Audio, i.e., the explanation was given in text and voice format simultaneously. The SA level factor had three levels: 1) SA L1, i.e., the explanation included only SA Level 1 information (i.e., perception), 2) SA L2, i.e., the explanation included SA Level 1 + Level 2 information (i.e., perception and comprehension), and 3) SA L3, i.e., the explanation included SA Level 1 + Level 2 + Level 3 information (i.e., perception, comprehension, and projection). Table cells represent the treated conditions in the experiment.

SA Level	Modality	
	Visual	Visual + Audio
SA L1	Text SA L1	Text + audio SA L1
SA L2	Text SA L2	Text + audio SA L2
SA L3	Text SA L3	Text + audio SA L3

A control condition was included in the experiment where participants did not receive any explanation.

5.2.2 Apparatus

The study was conducted using a survey developed in Qualtrics (Provo, UT) and was published in MTurk. The survey was designed to evaluate the effects of SA and explanation modality on participants' situational trust, explanation satisfaction, and mental workload in uncertain situations while driving an AV. The driving scenarios were presented in videos created in the CarMaker autonomous driving simulation environment (Karlsruhe, DE).

Table 5.2: Dependent variables

Measure	Description	Scale
Trust	Measured at the end of each scenario	STS-AD
Explanation Satisfaction	Measured at the end of each scenario	Explanation satisfaction scale
Mental Workload	Measured once participants watched all the 6 scenarios	DALI

5.2.3 Experimental Design

Independent variables. The experiment was a three (SA level: SA L1, SA L2, and SA L3) by two (modality: visual, visual + auditory) between-subjects factorial design with 6 scenarios. Alongside the 6 experimental conditions, a control condition with no explanations was also tested. The independent variables were the three levels of explanations mapped to three SA levels presented to the participants according to Endsley's SA model (*Endsley*, 1995a) and in two types of modalities, i.e., visual and visual + auditory. During the experiment, the participants' SA was measured through the Situation Awareness Global Assessment Technique (SAGAT) (*Endsley*, 1988). The SAGAT is a freeze-probe technique that requires pausing the simulation and asking a series of questions to assess the participants' awareness of the current situation. For each scenario, three different questions were developed to test the participants' perception of surrounding objects, comprehension of the current situation, and projection of the future state for that uncertain situation. All the questions designed for the SAGAT technique were developed based on a previous study (*van den Beukel and van der Voort*, 2017). Table 5.3 shows an example of multiple-choice questions for the training scenario (see Table 5.4). Regardless of the experiment conditions, for each scenario, three SA questions were included in the survey corresponding to three levels of SA. The participants obtained one point if they answered the question correctly. With three questions for each scenario, the participants could get as many as 18 points, indicating perfect SA.

5.2.4 Dependent Measures

The dependent variables in this study were situational trust, mental workload, and subjective satisfaction with explanations. Situational trust was measured by the self-reported Situational Trust Scale for Automated Driving (STS-AD) (*Holthausen et al.*, 2020a). The model evaluates situational trust in six categories: trust, perfor-

Table 5.3: Example questions for the training scenario to measure SA with a SAGAT Questionnaire.

Level of SA	Question	Options
Perception	The simulation just "froze". Which road user was in front of the AV?	1) Bus, 2) <u>Pedestrian</u>, 3) Cyclist, 4) I don't know, 5) Other
Comprehension	What caused you to seek your attention in this situation?	1) <u>Pedestrian's intention to cross the street</u>, 2) Approaching heavy traffic, 3) Approaching closed road, 4) Faulty road lanes, 5) I don't know, 6) Other
Projection	If the simulation resumes after this "freeze", what situation would require your extra attention or intervention?	1) Other road user's violations, 2) <u>AV's possibility to hit pedestrian</u>, 3) Impeding the traffic by stopping at intersection, 4) I don't know, 5) Other

* The underlined option indicates the correct answers.

mance, non-driving related task (NDRT), risk, judgment, and reaction, by asking the following questions: 1) I trusted the automation in this situation, 2) I would have performed better than the AV in this situation, 3) In this situation, the AV performed well enough for me to engage in other activities, 4) The situation was risky, 5) The AV made a safe judgment in this situation, and 6) The AV reacted appropriately to the environment. All the six STS-AD scales were measured with a 7-point Likert scale. Situational trust was measured right after the participant watched one video that depicted a specific driving scenario. Thus, it was measured six times for six scenarios.

To understand the subjective satisfaction of the given explanations, the explanation satisfaction scale developed by Hoffman et al. (2018) was used. In this study, it was presented to the participants with five items and was measured with a 7-point Likert scale. The following items were included: This explanation of how the AV behavior was 1) satisfying, 2) had sufficient details, 3) contained irrelevant details, 4) was helpful, 5) let me judge when I should trust and not trust the AV. Explanation satisfaction was also measured once right after the participant watched one specific
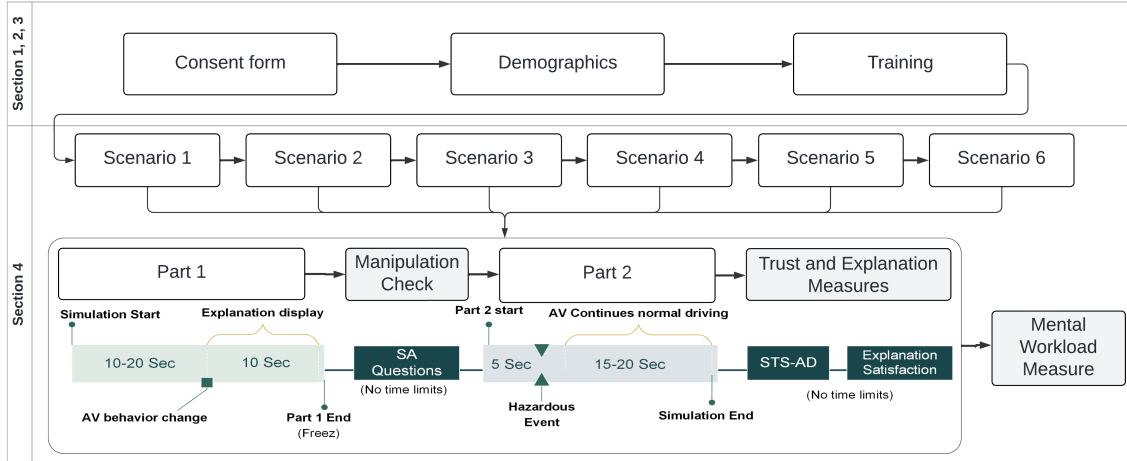
Figure 5.1: Survey procedure.

driving scenario. Thus, it was measured six times for six scenarios.

The mental workload was measured using the driving activity load index (DALI) (*Pauzié*, 2008), which is a revised version of the NASA-TLX and specifically adapted to the driving tasks. DALI includes six factors: attention, visual, auditory, temporal, interference, and stress. In order to reduce the time of taking the survey, the cognitive workload was only measured once at the end of the survey using a 7-point Likert scale when the participants watched all the six scenarios. In the control and text-only scenarios, the auditory demand was removed.

5.2.5 Survey Design and Procedure

The survey consisted of four sections as illustrated in Figure 5.1. The first section included a consent form. In the second section, the participants filled in a set of demographic questions. The third section was a training session, where the participants were given one simulation video example not used in the test session with three SA questions. Since the SA questions were designed based on the SAGAT technique, the freeze-probe technique was imitated for each scenario by dividing the simulation into two parts representing before and after the freeze situations. The fourth test section included six AV driving scenarios as shown in Table 5.4. The participants watched

58
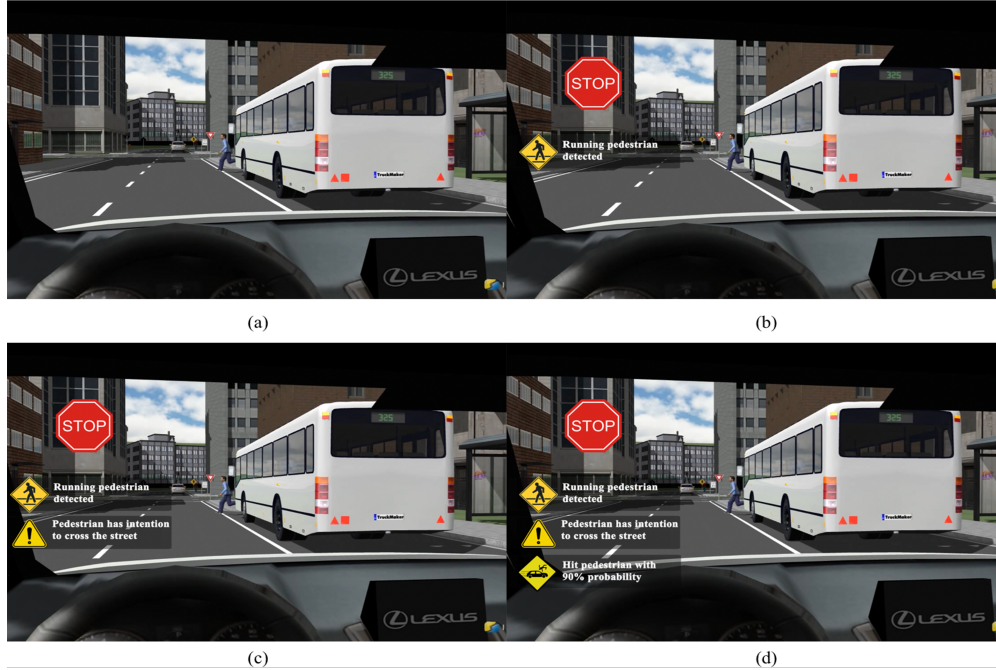
Figure 5.2: Presented explanations S2 in (a) control, (b) SA L1, (c) SA L2 and (d) SA L3 conditions (see S3 L3: https://youtu.be/GNL2cMK5Lyk).

the first part of each simulation video and answered three questions about their SA about the driving scenario (see Table 5.3). Then, they watched the second part of the video where they could see what happened actually. After each scenario, the participants evaluated their situational trust in AVs using the STS-AD scale and rated the given explanation(s) using the explanation satisfaction scale. After finishing all the six scenarios, the participants were required to report their mental workload about the explanations.

5.2.6 Scenario Design

Participants' trust in AVs' scenarios was investigated by manipulating their SA using three SA levels (*Endsley*, 1995a) in different scenarios. All the situations were extracted from real driving scenarios and from Wiegand et al.'s work (2020), where they explored the necessity of the explanations in unexpected situations while driving an AV. Seven scenarios were identified and simulation videos were created to visualize

Table 5.4: Scenarios with description in this study

Name	Scenario	Description and Link
Training	Reluctant to turn right due to a pedestrian	City: The AV stops before turning right, and a pedestrian stands on the other side of the street and moves a little. There is no crosswalk. The AV slowly turns with intermittent stopping. https://youtu.be/B3Zw7-kZzoY
S1	Long wait at the intersection to turn left	Highway: The AV approaches an intersection with a green traffic light. It stops behind the traffic light, and then moves a bit. After about 10 seconds, the AV finally turns left after an oncoming car passes. https://youtu.be/PfpsxPfmePg
S2	The AV stops and the pedestrian crosses	City: While driving, the AV stops abruptly. It waits. After seconds, a pedestrian crosses the street behind the bus. The AV continues driving. https://youtu.be/i9nt3FvqbnM
S3	Unexpected stop due to an emergency vehicle	City: The AV stops. In some distance, there is a green traffic light. After a while, an emergency vehicle passes with the siren on. The AV waits for about 2 more seconds and continues driving. https://youtu.be/XmSrxEYeySo
S4	Strong and abrupt braking to reach the speed limit	City: The AV enters the city and brakes abruptly and strongly to reach the speed limit. https://youtu.be/b5jrT4Mx9bg
S5	Early lane change due to heavy traffic	Highway: The AV changes to the right lane far away from the turn and it detects heavy traffic on the defined route. https://youtu.be/0kQw498WK20
S6	The AV waits for a long time before merging	Highway: The AV slows down and stops. It needs to merge with the highway and waits for its chance with a safe distance while the AV's intention in merging lanes is not clear. Traffic is overloaded. https://youtu.be/L8I8ULMcuYw

the situations (see Table 5.4). In each scenario, the corresponding information was embedded into the video explaining the current situation before the AV started its actions. In this work, explanation modality was also explored by adding voice-over to simulations. In visual+auditory conditions, an auditory message with a synthesized female voice was added to provide the same situational explanations simultaneously with the visual explanations. Figure 5.2 illustrates the simulations for the S2 scenario

(see Table 5.4) correspondingly for the control, SA L1, SA L2, and SA L3 conditions. In the control condition, no explanation was given. The SA L1 condition provided information explaining the perception of the current environment, including the surrounding objects which influenced on the AV's behavior. In the SA L2 condition, additional information was used to explain how the AV understood the surrounding environment. The SA L3 condition included all the information from SA L2 and added extra information about how that might affect the AV's behavior in the future.

5.2.7 Data Analysis

Statistical analysis was conducted using the R language in RStudio. A two-way ANOVA was used to analyze the effects of the explanations on situational trust, explanation satisfaction, and mental workload. The alpha was set at 0.05 for all the statistical tests. Post-hoc analysis was conducted with Tukey's HSD test.

5.3 Results

5.3.1 Manipulation Check

In this study, the effect of the provided information on SA was explored with the control condition and three SA levels, where the participant's SA was measured by the number of correct responses throughout the experiment. A two-way ANOVA test showed that there was a significant main effect of SA levels ($F(3, 333) = 38.23, p = .000, \eta^2 = .253$) and modalities ($F(1, 333) = 4.26, p = .040, \eta^2 = .009$) (see Figure 5.3). There was no significant interaction effect between SA levels and modalities ($F(2, 333) = 0.28, p = .752$). The post-hoc analysis showed that SA was significantly higher in SA L1, L2, and L3 conditions compared to the control condition, and significantly higher in the visual + auditory modality ($p = .040$) compared to the visual-only modality. Figure 5.3 illustrates the mean SA scores across different experimental conditions.
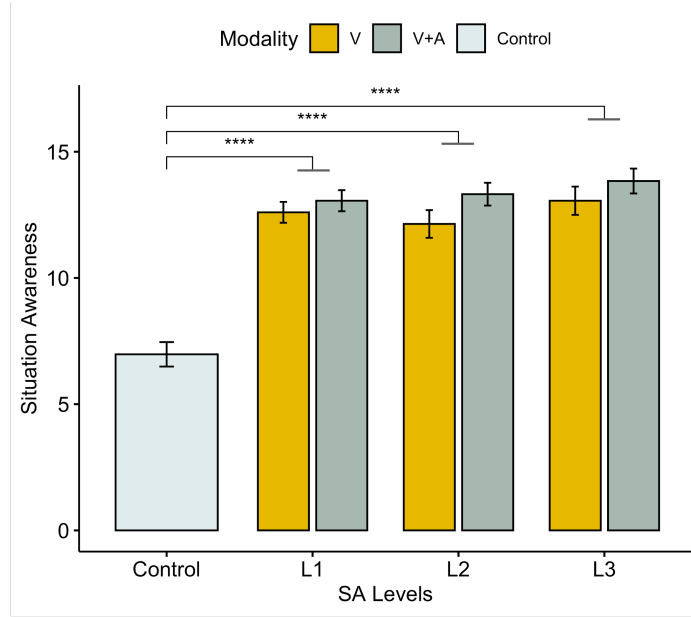
Figure 5.3: Mean SA scores at different conditions and explanation modalities with standard error, where '***' indicates $p < 0.001$.

5.3.2 Situational Trust

The means of the STS-AD over all six scenarios were calculated and analyzed with a two-way ANOVA. Results showed that the main effect of SA levels was significant $(F(2, 294) = 3.93, p = .020, \eta^2 = .029)$ whereas the main effect of modalities $(F(1, 294) = .07, p = .789, \eta^2 = .000)$ and the interaction effect $(F(2, 294) = 1.31, p = .272, \eta^2 = .007)$ were not significant (see Figure 5.4). The post-hoc analysis showed that STS-AD in SA L2 was significantly higher than in SA L1 $(p = .036)$. Specifically, STS-AD in Text SA L2 was significantly $(p = .040)$ higher than that in Text + Voice SA L1. And STS-AD was significantly higher $(p = .047)$ in SA L2 than that in SA L3. Specifically, STS-AD in Text SA L2 was marginally $(p = .052)$ higher than that in Text SA L3. Compared to the control condition, it was found that only SA L2 was significantly higher $(p = .011)$ mainly due to the visual-only modality $(p = .026)$. As for the visual + auditory modality, the difference was not significant $(p = .131)$.
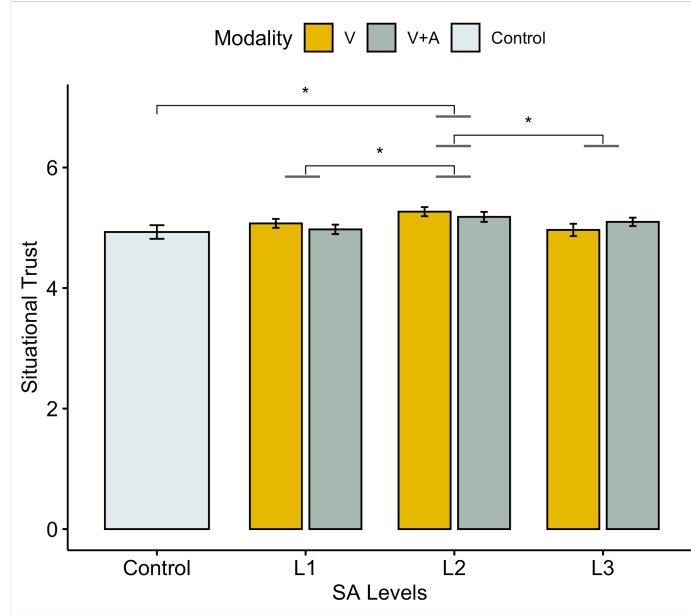
Figure 5.4: Overall mean and standard error of situational trust measured by the SA levels and modalities, where '*' indicates $p < 0.05$.

5.3.3 Explanation Satisfaction

With regard to explanation satisfaction, the two-way ANOVA showed a significant interaction effect ($F(2, 294) = 4.53, p = .012, \eta^2 = .030$). The post-hoc analysis showed that the participants were significantly more satisfied with the given explanations in the SA L1 ($p = .014$) and SA L2 ($p = .043$) conditions compared to the SA L3 condition when explanations were presented in the visual-only modality. Furthermore, in the SA L3 condition, when a comparatively large amount of explanation information was presented, a significant effect of explanation modality was found that the visual + auditory condition resulted in a higher satisfaction score compared to the visual-only ($p = .009$) condition (see Figure 5.5).

5.3.4 Mental Workload

The participants' self-reported mental workload was analyzed using the mean values of all the six DALI factors. As shown in Figure 5.6, we found a significant main effect of SA levels ($F(2, 294) = 3.70, p = .026, \eta^2 = .024$) that participants'
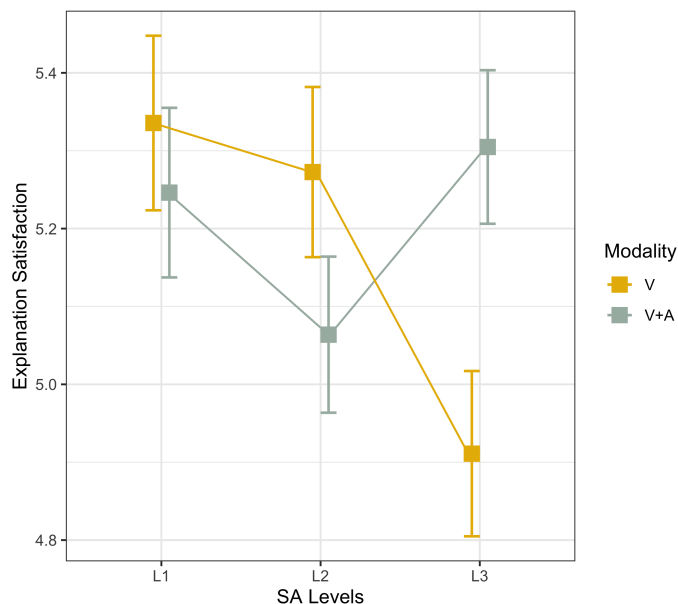
Figure 5.5: Interaction effect of SA levels and modalities with standard error on explanation satisfaction.

mental workload was significantly higher ($p = .018$) in the SA L2 condition than that in the SA L1 condition and than that in the control condition ($p = .009$). Specifically, we found that participants' mental workload in the Text SA L2 condition was significantly ($p = .016$) higher than that in the Text SA L1 condition and was significantly ($p = .012$) higher than that in the control condition. Thus, the significant differences were mainly caused by the visual-only modality.

5.4 Discussion

5.4.1 The Effects of SA

In this study, we investigated the effects of SA explanations and modalities on situational trust, explanation satisfaction, and mental workload in AVs. First, our results partially supported that SA levels positively affected participants' situational trust (see Figure 5.4) and SA L2 led to the highest level of situational trust. In this sense, situational trust appeared to be sensitive to SA. In particular, the par-
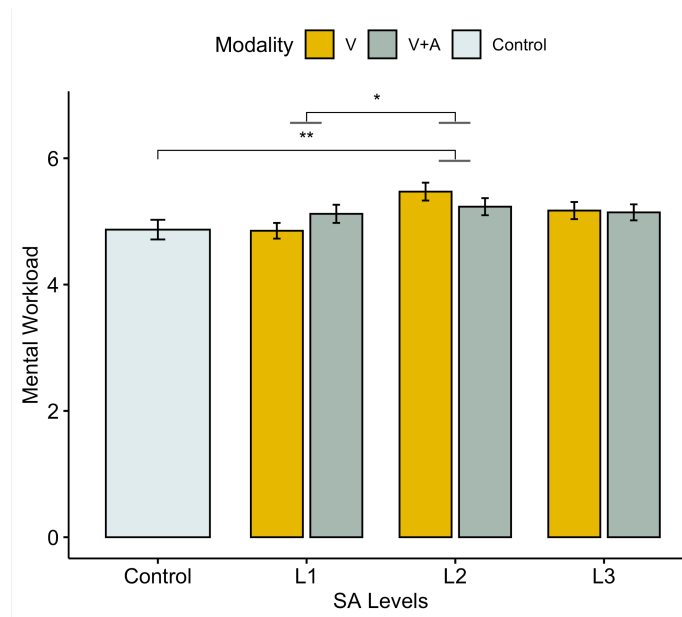
Figure 5.6: Overall mean and standard error of mental workload measured by the SA level and modality, where '*' indicates $p < 0.05$ and '**' indicates $p < 0.01$.

ticipants' trust was significantly higher in SA L2 compared to SA L1 and L3, where the given information was either too little to foster the participants' perception and comprehension of the current situation or was redundant to notably improve trust (*Mackay et al.*, 2019). One possible reason might be the out-of-the-loop problem, as Endsley et al. (1995) found that SA L2 was the most negatively affected level by automation, where people's understanding of the situation significantly decreased, pushing them out of the control loop. When SA L2 explanations were provided to help the participants understand the situations and bring them back to the control loop, their situational trust was significantly improved. Besides, consistent with Endsley (1995a), the participants might comprehend and project the future state at the same stage in SA L2, which indicates that the participants might already receive information that is supposed to receive in SA L3. For instance, in the scenario 2 (see Table 5.4) comparing the SA L2 explanation (i.e., L1: "Running pedestrian detected", L2: "Pedestrian has an intention to cross the street"), and SA L3 (i.e., L1, L2, and L3: "90% risk of hitting a pedestrian") explanations, the participants might project the

risk of accident at L2, hence the L3 explanation was not useful. Therefore, there was also no significant difference between SA L2 and SA L3 in terms of cognitive processing as shown in Figure 5.6.

With regard to the interaction effect of SA levels and modalities on explanation satisfaction (see Figure 5.5), the participants were more satisfied with the text explanations in SA L1 and L2 might be due to the machine-generated voice. As Tsimhoni, Green and Lai, (2001) showed that natural speech led to a better comprehension of the given information compared to synthesized speech. However, participants were more satisfied with the combined visual and auditory explanations in SA L3. This result was supported by the information processing theory (*Wickens*, 2008a) that it was easy to comprehend a large amount of information when more than one sensory resource (i.e., visual and auditory) was used while the participants might be annoyed to have redundant explanations with less information.

For cognitive workload, we found that participants had a higher cognitive workload in the SA L2 condition, especially the visual-only explanations, compared to the control and SA L1 conditions. One possible reason might be that the participants with explanations corresponding to SA L2 were actively interpreting the information to understand the driving scenarios, which improved their situational trust (see Figure 5.4). However, regardless of the extra information, SA L1 and SA L3 had similar levels of cognitive workload as the control group which might be due to the experiment design.

5.4.2 Implications

We proposed to explain AV behavior based on the three levels of SA and XAI theoretically to satisfy their informational needs in unexpected scenarios, and empirically explored its effects on human-AV interaction. Considering the AV as a black-box AI system, the properly-designed explanations based on the SA framework helped to

define which components in the system should be explained to meet drivers' informational needs in order to understand AV's behavior. While previous studies have focused on "how", "why" and "what" information for explanations empirically (*Koo et al.*, 2015, 2016; *Du et al.*, 2021), this SA-based model focused more on XAI concepts and reduced the complexity of the situations to understand how the AI system came to that particular decision systematically.

During the interaction between the driver and the AV, it is important that the AV provides explanations with different levels of SA for the driver to understand its decision-making process. As pointed out by *Sanneman and Shah* (2020), the key point is how to map such explanations into the needed three SA levels when designing such a black-box AV system as an XAI system. At SA Level 1, we need to provide explanations about what objects are perceived from the environment to explain the effects of external factors on the decision-making process. At SA Level 2, we should explain how the AV understands the situation by taking the perceived objects and their actions into consideration. At SA Level 3, we might consider what actions would the AV and other road users take in the near future. Our explanations attempted to be designed based on the theory-based SA model to satisfy drivers' informational needs and benefit them by improving their trust with a minimal level of cognitive workload.

5.4.3 Limitations and Future Work

This study also has limitations that can be examined in future studies. First, the experiment was conducted in a low-fidelity setting on MTurk due to the COVID-19 pandemic. The SA was measured with the SAGAT technique (*Endsley*, 1995a) and we found that participants' SA was notably improved compared to the control condition. However, we could not identify significant differences among the three SA levels based on the provided explanations. One of the possible reasons might be that the data

was collected on MTurk, where the scenarios were relatively short (30-45 seconds) and the fidelity was relatively low in the experiment. This potentially reduced the participants' engagement level. Another reason might be the absence of non-driving related tasks due to the difficulty in controlling participants when the experiment was conducted on MTurk, which allowed the participants to continuously monitor the ride. Nevertheless, the significant differences in SA between the control conditions and others indicated the importance of simple explanations in improving SA. Further investigations are needed to understand the effects of different explanations on SA and subsequently on trust, mental workload, explanation satisfaction, and the joint performance of the human-AV team in high-fidelity driving simulators. Second, only self-reported measures were used to evaluate the trust and mental workload. Additional measures, such as physiological measures (e.g., galvanic skin response (*Du et al.*, 2020a), eye-tracking (*de Winter et al.*, 2019)) can be included in future studies. Third, only a limited number of scenarios were tested in the experiment with low to moderate risks. Future studies can explore more scenarios with different levels of risk. Fourth, since the experiment was conducted as a between-subjects design, the participants experienced only one of the SA levels, the results might be affected by individual differences and the low-fidelity of the experiment setting.

5.5 Conclusion

In this study, we designed an SA-based explanation framework to help drivers understand the driving situations and map the AV's behavior properly to the situation. By exploring participants' situational trust, cognitive workload, and explanation satisfaction, we evaluated the effectiveness of the framework in three SA levels and two modalities. Based on the results, it was partially supported that SA-based explanations improved participants' situational trust. Among three levels, SA L2 resulted in higher situational trust and mental workload regardless of the explanation modality.

However, modality preferences were changed from visual to visual and audio due to the explanation amount in SA L3. Overall, the results confirmed that the properly-designed explanations based on the SA-based framework helped orient drivers in the unexpected situation and assess the AVs' behavior accurately leading to higher trust and acceptance of these vehicles.

Chapter 6 Enhancing Situation Awareness for Conventional Vehicles in Mixed Traffic Environment

6.1 Introduction

as AVs become more prevalent on the roads, they will coexist with conventional vehicles (CVs) for the foreseeable future, creating mixed traffic environments (*Juhlin*, 1999). This coexistence presents unique challenges, particularly in terms of communication between AVs and other road users, such as CVs and vulnerable road users (VRUs). In traditional vehicle interactions, human drivers or VRUs rely on a limited set of explicit signals such as horns, headlight flashes, and eye contact or gestures as well as implicit cues like vehicle speed change and trajectory (*Miller et al.*, 2022; *Dietrich et al.*, 2019a). These communication methods, while limited, are well-understood by humans who can interpret and respond to them based on shared understanding and driving experience. However, in AV-human interactions, these traditional methods become less effective or even obsolete. AVs lack the ability to make eye contact or use gestures, and their movements may not convey the same implicit intentions as those of human-driven vehicles (*Papakostopoulos et al.*, 2021; *Brown et al.*, 2023).

This lack of communication can lead to misunderstandings, confusion, and potentially hazardous situations, especially in complex traffic scenarios like intersections or merging lanes (*Brown et al.*, 2023). In such situations, both vehicles influence and respond to each other's actions, requiring quick response based on communication and

70

mutual understanding. CV drivers may exhibit complex and nuanced behaviors that are not always rational or predictable, making it challenging for AVs to anticipate and respond effectively to these interactions (*Siebinga et al.*, 2023; *Isele et al.*, 2018). To ensure smooth traffic flow and minimize the risk of accidents, AVs need to be able to engage in these interactions in a safe, efficient, and socially acceptable manner. Therefore, there is a need for new, explicit communication methods that bridge the gap between AVs and CVs, ensuring that human drivers can understand and anticipate AV behaviors as intuitively as they do with other human drivers (*Schieben et al.*, 2019).

To address communication challenges between AVs and CVs, researchers have investigated vehicle-to-everything (V2X) communication systems, which enable real-time exchange of vehicle-specific information (e.g., speed, position, and intended maneuvers) and environmental data (e.g., road hazards and emergency vehicle locations) (*Kenney*, 2011). These advanced systems extend the capabilities of traditional visual cues, minimizing potential misinterpretations between human drivers and AVs. The resulting shared awareness promotes a more predictable and orderly traffic environment, enhancing situational awareness for all road users (*Shladover*, 2018). However, realizing these benefits hinges on the development of a standardized communication framework, which must be accepted and implemented across diverse vehicle types and manufacturers to ensure successful cooperation of AV and CV systems (*Guanetti et al.*, 2018). Another important communication strategy is to explicitly display the intention on an external human-machine interface (eHMI). so that other road users can understand AV's intention easily. In both cases, the design of the communication system must consider the human driver's mental model and informational needs to establish effective communication and avoid causing additional cognitive disturbance. This aspect remains less investigated in the AV-CV context. In contrast, numerous studies have explored these human factors in AV-VRU interactions through the devel-

opment of eHMIs. By utilizing visual cues, such as LED displays, symbols, and laser projections, as well as auditory signals like beeps and chimes, studies have shown that eHMIs can significantly enhance traffic safety by explicitly conveying AV's intentions (*Bai et al.*, 2021; *Palmeiro et al.*, 2018; *Rasouli and Tsotsos*, 2019). Research has demonstrated that in contrast to implicit cues derived from an AV's trajectory or speed changes (*Fuest et al.*, 2018), eHMIs can reduce the time pedestrians take to cross the street and improve the safety of their crossing decisions (*Dietrich et al.*, 2019b). Furthermore, eHMIs have been found to increase the perceived trustworthiness and intuitiveness of AVs (*Hensch et al.*, 2020). These findings highlight the potential of eHMIs in enhancing traffic safety by reducing misunderstandings and facilitating decision-making processes during human-AV interactions.

In this study, we aimed to address the communication need by investigating how HMIs can facilitate effective communication between AVs and CVs in mixed traffic environments, particularly in challenging traffic situations. We employed a within-subjects experimental design in a VR environment. The independent variable was the HMI design, which had three distinct levels. The first level served as a control condition without any HMI implementation. The second level involved an external HMI (eHMI) positioned on the front of the AV. The third level featured an internal HMI (iHMI) integrated within the CV. To gain deeper insights into the HMI's efficacy, we introduced two patterns within the eHMI and iHMI conditions that were developed using our SA framework. One signaling the AV's intention to yield, and the other indicating its priority right of way. Our assessments utilized both self-reported and objective measures. We evaluated the effect of these design variations on drivers' SA, trust in AV, acceptance of HMI deployments, and the cognitive workload during interaction with the AV.

6.2   Method

6.2.1   Scenario

To investigate communication challenges in high-risk traffic situations, we selected the 'Left Turn across Path from Opposite Directions at Signalized Junctions' scenario based on Najm et al.'s crash report (2007), which identified this scenario as a significant contributor to crashes, injuries, and fatalities. Although the right-of-way rules in such scenarios are generally understood (straight traffic has priority over turning traffic), the malfunctioning traffic lights introduce a critical uncertainty: how will the AV, as a non-human road user with potentially unfamiliar behavior patterns, navigate this situation (see Fig. 6.6). The scenario presents a unique opportunity to explore the communication gap between AVs and CVs. Will the CV driver accurately anticipate the AV's actions, or will the lack of explicit communication lead to hesitation, misinterpretation, or even collisions? By examining this scenario, we aim to uncover the potential benefits and limitations of different communication strategies between AVs and CVs in resolving such uncertainties and ensuring safe navigation in mixed traffic environments.

6.2.2   Designing HMIs with a Human-centered Approach

In the first phase of this study, we employed a human-centered design process (*IDEO*, 2015) to develop interfaces to facilitate communication between CVs and AVs for a specific scenario in intersections. To empathize with CV drivers and understand their needs, we reviewed existing literature and had a class discussion (with more than 30 master students in the human-centered design program in University of Michigan-Dearborn) about the challenges faced by communications between CVs and AVs. Based on our analysis, we identified three main challenges, including: 1) conspicuousness - easy to see visually, 2) comprehensibility - easily understandable

with minimal cognitive effort, and 3) identifiability - easy to recognize to whom it was addressed.

Then, we had a brainstorm session to generate a wide range of possible ideas for the "left turn" scenario in an intersection (see Table 6.4) and followed the design principles outlined in Rettenmaier et al.'s study (*Rettenmaier et al.*, 2020) and Avetisyan et al.'s SA framework (*Avetisyan et al.*, 2022) for informational needs. Ultimately, we narrowed down to eight design concepts (see Fig. 6.1) that included two versions for the AV to 1) yield and 2) insist on the right of way, and were presented using different visual formats, i.e., signs, texts, or a combination of both. Designs 1 to 4 included SA level 1 (i.e., perception of the items in the environment,) and level 2 (i.e., comprehension of the current situation) information, while designs 5 to 8 included additional SA level 3 information (i.e., projection of future status of the environment). To facilitate communication between CVs and AVs in the selected scenario, we broke down the HMI message into three parts, which explained the traffic situation, a suggested way of behaving, and improved trust. Firstly, we added a sign or text that described the current issue, stating that the traffic lights were not functioning properly. Secondly, we included right-of-way information that informed CV drivers how to proceed, incorporating well-known traffic signs and colors to enhance comprehension. Thirdly, we attempted to provide extra information to increase the confidence of CV drivers.

To evaluate the interface designs, we conducted an online survey study with 32 participants who were introduced to the "left turn" scenario and asked to show their level of agreement for the prototypes of the design concepts using a 7-point Likert scale based on four statements: 1) The message is easy to understand, 2) The message contains relevant details, 3) The message helps to respond quickly, and 4) The message is preferred with text only. The final score was calculated as the average of these four statements. The analysis of the online survey data revealed a significant difference among the eight concepts ($F(7,8) = 10.05, p < 0.000$). Notably, the second interface
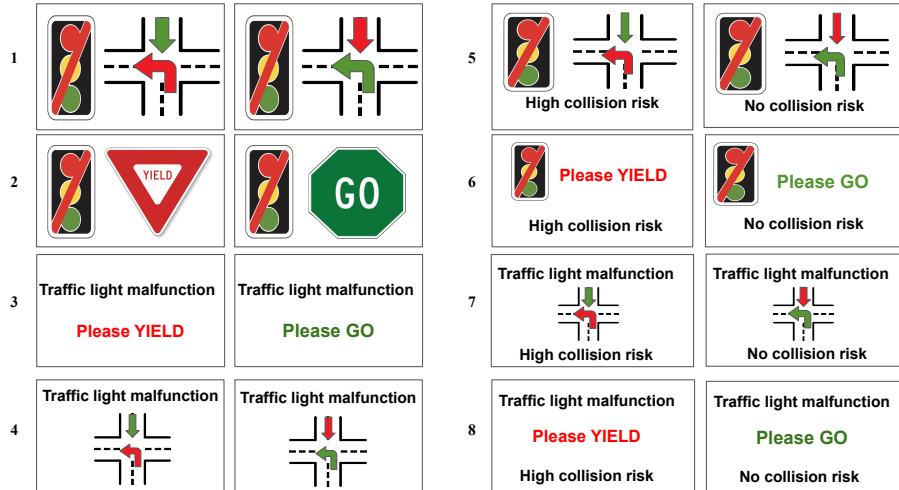
Figure 6.1: Interface design concepts for AV-CV communication. In all the concepts, the first figure showed that traffic lights are malfunctioning and CV had to yield the AV, while the second figure showed that CV had right of way.

design (see Fig. 6.1) received the highest rating of 5.28 and was chosen to be tested in the second phase of the study, which involved using a driving simulator in a VR environment.

6.2.3 Participants

This study was conducted in accordance with the ethical requirements of the Institutional Review Board at the University of Michigan (application number HUM00219554). Consent forms were sent in advance, allowing thorough review before confirming participation. In the design phase, 32 students participated in evaluating eight concepts. These participants were recruited from a Human-centered Design Engineering course focusing on user-centered design principles and advanced technologies, including autonomous vehicles. Our recruitment process emphasized the voluntary nature of participation and its independence from academic performance. All enrolled students were eligible to participate, with no predetermined limit on participant numbers. As an incentive, students were offered a minor grade bonus, equivalent to 1% of their course score, for their participation.

In the experimental phase, a total number of 50 participants took part in the experiment. Due to the severe motion sickness and eye calibration failures, six participants could not finish the experiment. Therefore, the data analysis was conducted based on the remaining 44 participants (11 females and 33 males; Age = 24.4 ± 4.19 years old) who were university students located in the United States and possessed a valid U.S. driver's license. On average, participants had 5 ± 4.52 years of driving experience and drove approximately 5 ± 1 days per week. Participants received compensation of $20 in cash upon completion of the study. The average completion time was 35 minutes for male participants and 41 minutes for female participants. During the experiment, participants were allowed to withdraw at any time and receive $5 per 10 minutes. Participants experienced low levels of motion sickness after the experiment(i.e. 1.66 for male and 1.70 for female participants).

6.2.4 Apparatus

The experiment was conducted in Virtual Reality driving simulator at the University of Michigan-Dearborn using a desktop computer with an Intel Xeon(R) W-2104 CPU processor running at 3.20GHz, 64.0 GB of RAM, and an NVIDIA GeForce RTX 3060 graphics card with 12 GB of memory. The operating system used was Windows 10. The experiment employed an HTC Vive Pro Eye headset (Taipei, Taiwan) in combination with the Logitech G29 Driving Force (Lausanne, Switzerland). Four different drives (one for training and three for experimentation) were created using the Unity game engine (San Francisco, CA). All self-reported data was collected using a survey developed and administered through the Qualtrics (Seattle, WA) platform.

6.2.5 Experiment Design

In the second phase of this study, the final design concept was evaluated in the VR driving simulator experiment.

Figure 6.2: Experiment setup with VR driving simulator. Participant was wearing HTC Vive headset and driving a CV using Logitech G29 steering wheel and pedals. The screen shows the participant view in the VR environment.

Independent variables. In this study, a within-subjects design experiment was conducted in which the independent variable was the communication interface condition with three levels: control, eHMI, and iHMI. The control condition did not have any explicit communication between the AV and the CV. In the eHMI condition, the AV communicated with the CV through an external display attached to front of the AV, which shared the current traffic situation and right-of-way information from the egocentric perspective (the CV driver's perspective). In the iHMI condition, the AV communicated with the CV through an internal display that appeared on the CV's heads-up display and shared the same information as described for eHMI.

Dependent variables. The study collected both self-reported and objective data (i.e., eye tracking and vehicle dynamics). Self-reported measures were used to assess SA, trust in AV, interface acceptance. *Situation awareness* was assessed using

a modified version of Situation Awareness Global Assessment Technique (SAGAT) technique (*Endsley*, 1995b). To maintain participants' engagement in VR and minimize motion sickness, we modified the original SAGAT method by conducting SA evaluations post-trial instead of interrupting the trials. Following previous studies (*Avetisyan et al.*, 2022; *van den Beukel and van der Voort*, 2017), we developed SA questions that aligned with SAGAT's principles and includes all three levels of SA: perception (Level 1), comprehension (Level 2), and projection (Level 3) (*Endsley*, 1988). After each trial, the participants were asked to answer SA questions (see Table 6.1) for two interactions separately, where they chose all the applicable answers out of the 5 possible options. The final SA score was measured based on the number of correct answers, with a score range of 0 to 3. To prevent negative impacts on participants' engagement in VR and to avoid causing additional motion sickness, the SA was evaluated in the post-trial phase. *Trust* in AV was measured with Jayaraman et al.'s (2019) version of Muir et al.'s (1987) trust scale with 7-point Likert scales. At the end of each session, participants evaluated their trust in five dimensions: competence, predictability, dependability, responsibility, reliability over time and faith, by answering question presented in Table 6.2. To understand the acceptance of the proposed concepts as communication interfaces, Van der Laan et al.'s (1997) nine-item acceptance measure was applied, where participants rated them in two dimensions: 1) perceived usefulness (i.e., useful, good, effective, assisting, and raising alertness), which focused on the functional aspects of the concepts and how well it could assist the participant in current situations, and 2) satisfaction (i.e., pleasant, nice, likable, and desirable) which focused on the overall emotional responses and fulfillment of expectations after experiencing the concepts. The measures were assessed by 7-point rating scales (see Fig. 6.4) after each trial was completed. Furthermore, the participants' simulation sickness was evaluated using the Simulator Sickness Questionnaire (SSQ) (*Kennedy et al.*, 1993) on a 7-point Likert scale.

To collect eye-tracking data from participants, the integrated eye tracker in the HTC VIVE Pro Eye headset was used with a resolution of 1440 x 1600 pixels per eye, and a 110-degree horizontal field of view. The eye tracker had a 120 Hz frequency of gaze data output with an accuracy of 0.5-1.1 degrees. In total, 7 measures were collected and used to analyze the pupil diameter and eye openness (see Table 6.3). Previous studies showed that mean change in pupil diameter was a reliable measure since it can eliminate the side factors, i.e. environment illumination, that could potentially influence on pupil diameters (*Palinko and Kun*, 2012). Therefore, to investigate how participants' mental workload was influenced by the interface condition, mean pupil diameter and eye openness changes were examined. Prior to the analysis, the raw data went through four step of data processing. First, the invalid eye data rows and outliers were removed. Next, the records of left and right eyes were combined by their mean coordinate values. Third, the interaction moments were identified using vehicle positions in the VR and the data was segmented to ten second before and after the interaction with AV. Finally, the mean pupil diameter change and eye openness change were calculated per interaction, and the average of two interactions was used for further analysis.
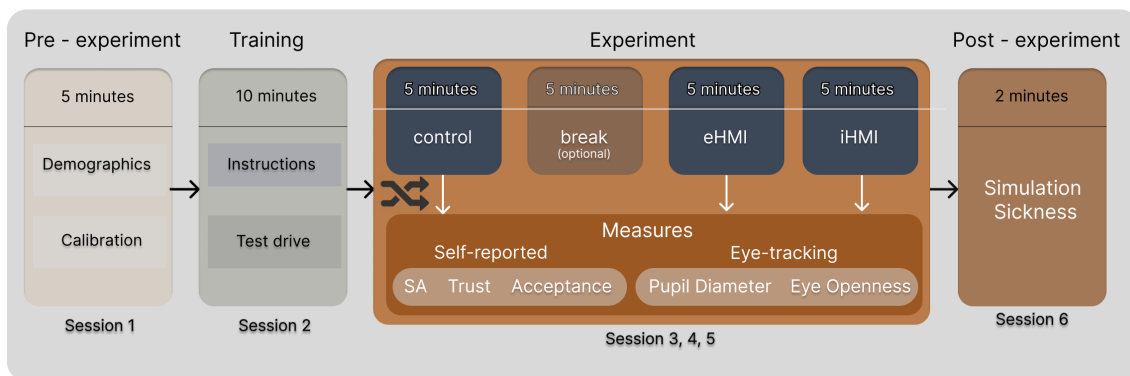
Figure 6.3: Experiment procedure.

79

Table 6.1: An example question for the iHMI condition to measure SA with a SAGAT questionnaire (*Endsley*, 1995a)

Question

You and the incoming automated vehicle approached an intersection as shown in the picture. Based on your understanding, what was the situation at the intersection (pick all applicable)?

Options

1. The traffic lights are malfunctioning

2. The incoming automated vehicle will yield

3. The incoming automated vehicle has the right of way

4. The incoming automated vehicle would expect you to yield

5. The incoming automated vehicle violated traffic rules

Note: The underlined options indicate the correct answers.

6.2.6 Experiment Procedure and Data Collection

The experiment consisted of six sections. In the first section, participants completed the demographic section of the survey and received an introduction to the experimental process and tools. Following the introduction, the experimenter calibrated the eye tracker, and the participants had a training session where they had an opportunity to familiarize themselves with the VR environment and driving equipment

Table 6.2: The trust questionnaire (*Jayaraman et al.*, 2019).

Dimension	Question
Competence	To what extent did the autonomous cars perform their function properly i.e., recognizing you and reacting for you?
Predictability	To what extent were you able to predict the behavior of the autonomous cars from moment to moment?
Dependability	To what extent can you count on the autonomous cars to do its job?
Responsibility	To what extent the autonomous cars seemed to be wary of their surroundings?
Reliability over time	To what extent do you think the autonomous car's actions were consistent through out the interaction?
Faith	What degree of faith do you have that the autonomous cars will be able to cope with all uncertain ties in the future?

Figure 6.4: The acceptance scale (*Van Der Laan et al.*, 1997). The usefulness measure was the average of useful, good, effective, assisting, and raising alertness items. The satisfaction measure was the average of pleasant, nice, likable, and desirable items.

through a test drive (see Fig. 6.2). During this training session, participants experienced interactions with AVs similar to the experimental scenarios and were introduced to two interface concepts. Upon completion of the training session, participants were given the choice to either continue or stop their participation in the experiment. Since a within-subjects experimental design was employed, after the training sessions, par-

Table 6.3: Eye tracking measures. Each measure was collected for the left and right eye separately.

Measures	Description
Timestamp	The current time of data recording
Eye validity	The bits explaining the validity of eye data
Eye openness	The level of eye openness
Pupil diameter	The diameter of pupil

ticipants experienced each of the three experiment sessions corresponding to three conditions, i.e., eHMI, iHMI, and control conditions, in a randomized order. At the beginning of each session, the participants received brief introduction about current interface. During the drive, the eye tracking outputs and vehicle trajectory were recorded. At the end of each session, the simulation was paused and the participants filled in the survey section required to measure the dependant variables (i.e., SA, trust, and acceptance) as showed in Fig. 6.3. At the end of the last session, participants were asked to evaluate the 16 symptoms of simulation sickness listed on the SSQ. Overall, the experiment took approximately 38 minutes to complete, with 15 minutes allotted for demographic information, instruction, and training, and approximately 5 minutes for each experimental session. The participants also had the option to take a 5-minute break between each experiment condition.

6.2.7 Scenario Design in VR

In the VR setup, the scenario was implemented as follows: Each condition featured two variations, with an AV situated at two different intersections. Instructed by the integrated navigation system's voice, the CV driver was directed to turn left (as shown in Fig. 6.6). In one interaction, the AV yielded its right of way, while in the other interaction, it insisted on the right of way. When the distance between the AV and the CV reached the predefined distance, the AV displayed a message informing that the traffic lights were not functioning, and the CV either had the right of way or had

Table 6.4: Vehicles' interaction scenario (*Najm et al.*, 2007)

Scenario	Factors	Severity
A conventional vehicle (CV) is making a left turn at an intersection in an urban area during daylight hours, with clear weather conditions and a posted speed limit of 30 mph. Although the intersection is signalized, the traffic lights are currently malfunctioning. At the same time, an autonomous vehicle (AV) is approaching the intersection from the opposite direction and proceeding straight through. The CV driver must navigate the intersection by crossing in front of the AV's path, due to the left turn (see Figure 6.6).	Intersection, low speed road, vision issues, situation unawareness, vulnerable driver are over-represented.	According to the crash report, this scenario was responsible for 190,000 crashes (3.19% of all crashes). These crashes involved a total of 389,000 vehicles (3.64% of all vehicles) and 558,000 people (3.71% of all people involved in crashes). Of the people involved, 1.24% suffered high-level MAIS 3+ injuries (i.e., serious, severe, critical, or fatal).

to yield to the AV.

In the eHMI condition, the messages were displayed on the front of the AVs as shown in Fig. 6.5c and 6.5d, disappeared as the vehicles passed each other. In the iHMI condition, the message was displayed on the CV's windshield as shown in Fig. 6.5a and 6.5b. In the control scenario, the AV did not have a communication interface, and the CV driver had to rely on their own understanding to navigate.

6.3 Results

As the experiment was conducted using a within-subjects experimental design, a one-way repeated measures ANOVA was conducted to examine how the interface conditions affected the dependent variables. For the post-hoc analysis, a Bonferroni correction method was applied to determine significant differences between different conditions. The statistical analyses were carried out using MATLAB and R pro-

(a)

(b)

(c)

(d)
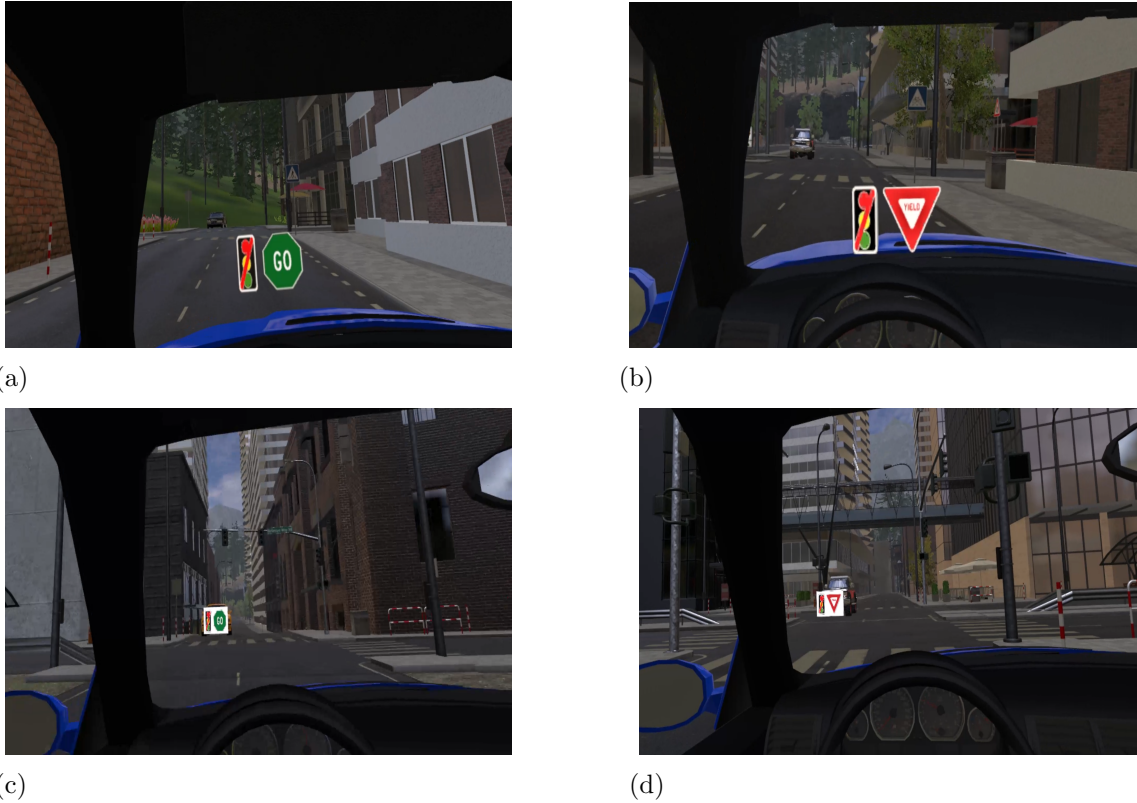
Figure 6.5: Experimental scenarios with (a) go and (b) yield messages in iHMI condition, (c) and (d) in eHMI condition

gramming languages, with an alpha level of .05 set for all the tests. The effects of the interface conditions on the dependent variables were assessed using η^2, which measures the proportion of variance in the dependent variables that can be accounted for.

6.3.1 Situation Awareness

SA was measured at two interactions for each interface condition and the mean SA score was used in the analysis. The results of one-way repeated measures ANOVA showed that there was a significant difference among three conditions ($F(2, 131) = 13.64, p = .000, \eta^2 = .132$). As illustrated in Fig. 6.7, the post-hoc analysis indicated that sharing the information through eHMI ($p = .002$) and iHMI ($p = .000$) significantly increased the SA level of the driver in the CV compared to the control
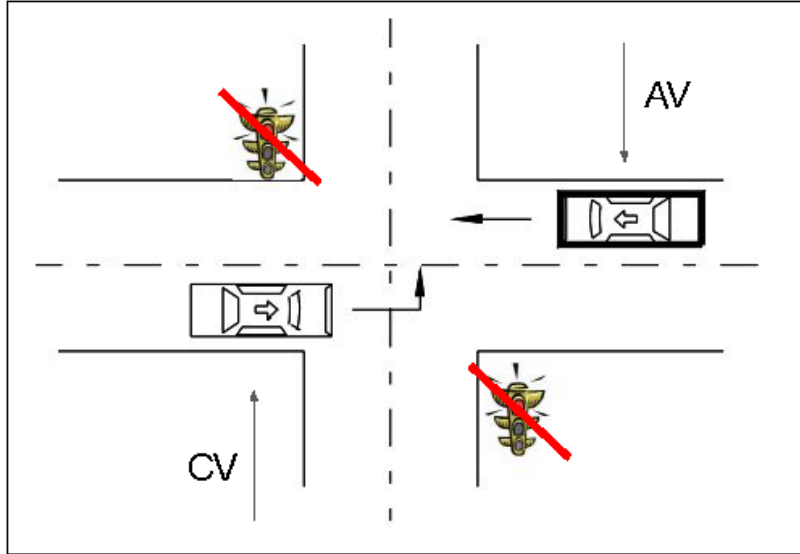
Figure 6.6: Conventional vehicle is turning left at non-signalized intersection and has to cross the path of incoming autonomous vehicle. The scenario was tested in two different AV behaviors at two different intersections. In one scenario, the AV yielded to the CV, while in the other, the AV insisted on the right of way.

condition. The difference between iHMI and eMHI was not significant.

6.3.2 Trust

In order to understand the effects of three interface conditions on trust, the mean of the five trust dimensions was analyzed. The results indicated that trust was significantly different among the tested conditions ($F(2, 131) = 25.20, p = .000, \eta^2 = .233$). Specifically, the trust level was significantly lower in the control condition than that in the eHMI and iHMI conditions ($p = .000$) as shown in Fig. 6.8. Also, in the eHMI condition the reported trust level was lower than iHMI, but the difference was not statistically significant ($p = .648$).

6.3.3 Acceptance

Regarding the acceptance of designed interface concepts, the results showed that participants' ratings were significantly different across the tested conditions with regard to usefulness ($F(2, 131) = 80.84, p = .000, \eta^2 = .535$) and satisfaction ($F(2, 131,) =$
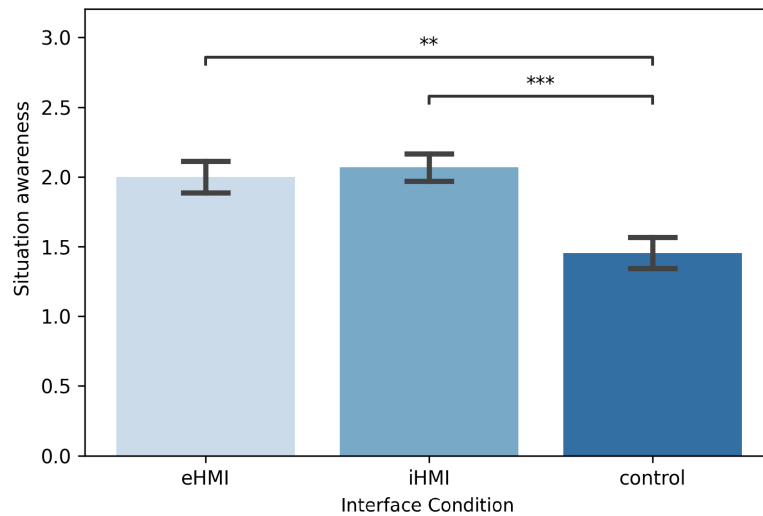
Figure 6.7: The effect of interface conditions on the SA level of the driver in the CV, where '**' indicates $p < .01$ and '***' indicates $p < .001$. Note that the error bar showed the standard deviation.

$45.80, p = .000, \eta^2 = .360$) of the concepts. Pairwise comparisons indicated that participants rated the interface as significantly useful compared to AVs without explicit communication interface. Additionally, the individual item comparisons in the usefulness dimension showed that there was a significant difference in the interface effectiveness where the iHMI condition was significantly higher compared to eHMI condition ($p = .015$). As for satisfaction items, no significant difference was found between eHMI and iHMI conditions ($p = 1.000$).

6.3.4 Pupil Diameter and Eye Openness

Due to the eye tracker's technical issues, 6 participants data were partially missing and were excluded from eye-tracking results. Therefore, eye-tracking measures were analyzed based on data from 38 participants.

The results of one-way ANOVA performed on mean pupil diameter change showed that there was a significant difference among three conditions ($F(2, 113) = 6.69, p = .002, \eta^2 = .185$). The post-hoc test showed that pupil diameter change in iHMI
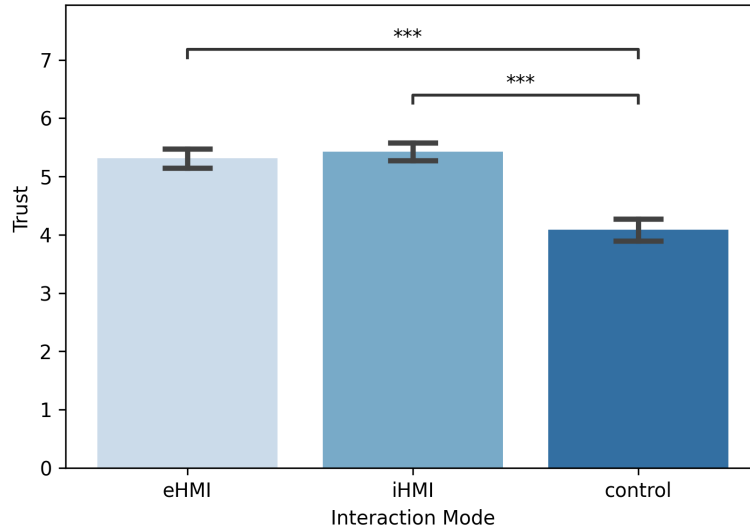
Figure 6.8: The effect of interface conditions on trust in AVs, where '***' indicates $p < .001$. Note that the error bar showed the standard deviation.

condition was significantly lower than that in control ($p = .001$) and eHMI ($p = .042$) conditions (see Fig. 6.10). Regarding the mean eye openness change, the patterns showed that in the control condition, participants tended to frequently squeeze their eyes during the interaction compared to the eHMI and iHMI conditions (see Fig. 6.11), however, the results of ANOVA showed that this difference was not statistically significant ($F(2, 113) = 1.40, p = .253, \eta^2 = .057$).

Regarding the mean eye openness change, the patterns showed that in the control condition participants tend to frequently squeeze their eyes during the interaction compared to the eHMI and iHMI conditions (see Figure 6.11), however, the results of ANOVA showed that this difference was statistically not significant ($F(2, 59) = 1.40, p = .253, \eta^2 = .057$). The η^2 for this study was .057, indicating a small effect size.

6.3.5 CV Speed

The results of a one-way ANOVA conducted on the mean speed change revealed a significant difference among the three conditions ($F(2, 59) = 5.59, p = .006$) when
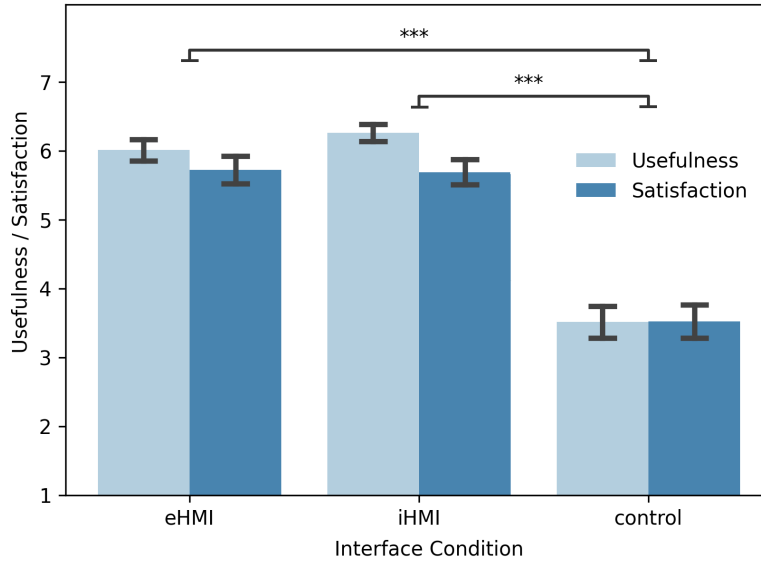
Figure 6.9: Perceived usefulness and satisfaction of HMIs, where '***' indicates $p <$ .001. Note that the error bar showed the standard deviation.

the AV asserted its right of way. According to the post-hoc test, the iHMI condition showed a significant drop in the CV speed after displaying the "Yiel" message compared to the eHMI condition ($p = .005$) and marginal drop compared to the control ($p = .096$) conditions (see Fig. 6.12). However, in the intersection where the AV yielded the right of way, the speed change was not statistically significant ($F(2, 59) = 1.32, p = .548$). Comparing the average CV speed, the results showed that in control condition participants' speed was significantly lower than other conditions regardless of the intersection (see Fig. 6.13). Specifically, at the intersection where CV had the right of way to turn, the participants approached the intersection with significantly lower speed than in iHMI condition ($p = .005$). Similarly, at the "Yield" intersection, their speed was significantly lower than that in the iHMI ($p = .027$) and eHMI ($p = .000$) conditions.
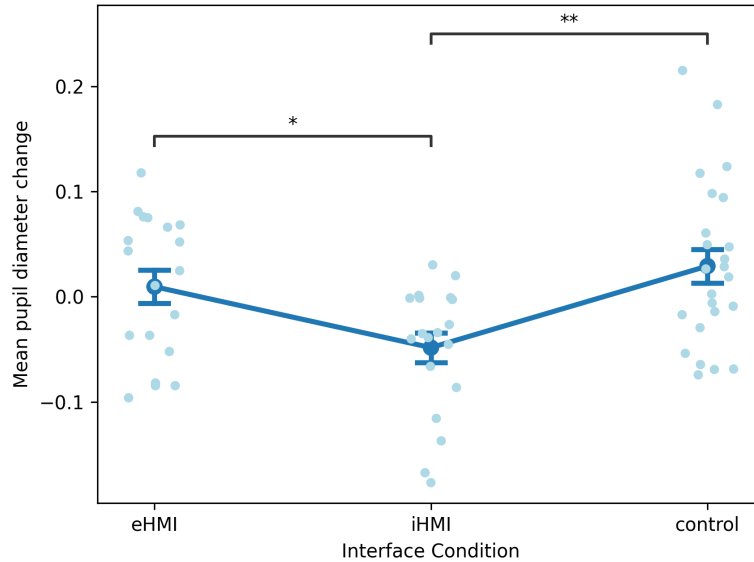
Figure 6.10: Mean pupil diameter change and standard deviation across different interface conditions, where '*' indicates $p < .05$,'**' indicates $p < .01$. Note that a positive change indicated larger pupil diameter following the interaction between CV drivers and AV, while a negative change indicated smaller pupil diameter.

6.4 Discussions

In this study, we developed and investigates the impact of HMIs as communication method between CV and AVs in mixed traffic environments. Specifically, we tested two distinct HMIs: external (eHMI) and internal (iHMI), in addition to a control interface (i.e., without any form of HMI communication), and evaluated their impact on CV drivers' SA, trust, acceptance, and mental workload.

Our results demonstrated that both eHMI and iHMI significantly enhanced CV drivers' SA compared to the condition without any explicit communication (see Fig. 5.3). With the HMI conditions, participants exhibited an increased consciousness of the traffic situation. In particular, they were able to promptly identify what was the traffic issue at the moment and and determine who had the right of way based on the shared information, enabling quicker responses in vehicle control. In contrast, during the control condition, drivers tended to maintain a safe distance until they could discern the AV's intention from implicit cues, such as speed changes or continuous
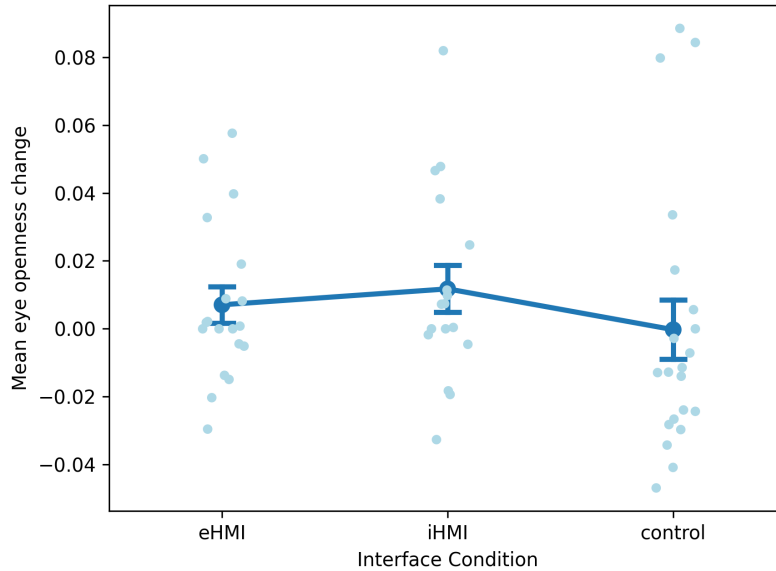
Figure 6.11: Mean eye openness change and standard deviation across different interface conditions. Note that a positive change in eye openness indicated wider eyes following the interaction between CV drivers and AV, while a negative change indicated narrower eyes.

driving. Furthermore, participants drove at significantly slower speeds in the control condition (see Fig.6.13), which could be attributed to the ambiguity of the situation. This observation confirms the notion that HMIs have the potential to bridge the communication gap between AVs and CVs, thereby reducing uncertainty and improving traffic flow efficiency with more seamless interactions. These findings aligns with previous work (e.g., (*Papakostopoulos et al.*, 2021; *Fuest et al.*, 2018; *Dietrich et al.*, 2019b)) that has demonstrated the benefits of explicit AV communication for VRUs. However, our study extends these insights to CV drivers, confirming that explicit communication is effective not only for VRUs but also for human drivers navigating mixed traffic scenarios, specifically at uncontrolled intersections.

The observed difference in trust towards AVs with HMIs compared to no communication highlighted the importance of explicit interfaces in enhancing the perceived reliability and dependability of AVs. Additionally, a notable behavioral difference was observed among the conditions. In the control condition, participants were explicitly
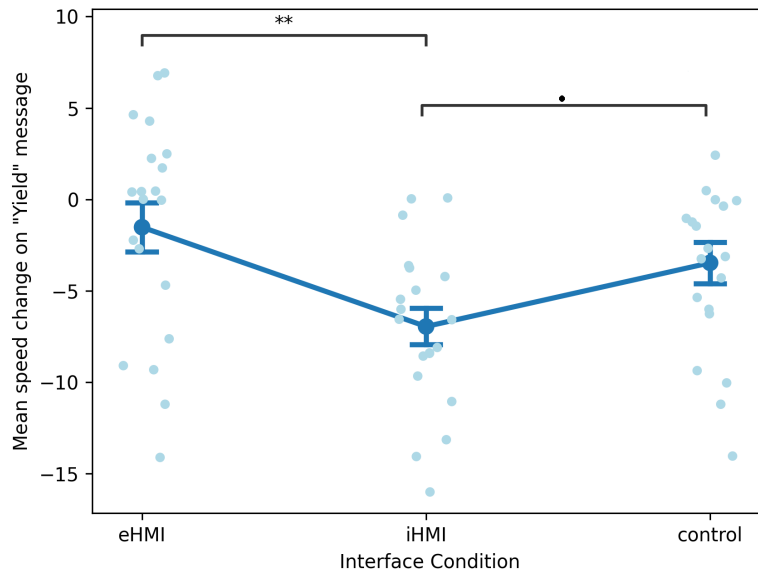
Figure 6.12: Mean speed change and standard error across different interface conditions. The change was measured by calculating the average speed three seconds before and after receiving the message. Note that a positive change in speed indicated the CV had positive acceleration after onset of "Yield" message, while a negative change positive braking after the message.

informed that the AVs would not communicate with them. As a result, they tended to maintain a lower speed in comparison to the HMI conditions. Despite this, their reaction to yield the AV was still delayed as they took more time to decelerate their vehicles compared to the time required for the AVs to complete their intersection crossing (see Fig. 6.13). As for the eHMI and iHMI conditions, the participants were more inclined to confidently rely on the instructions in the iHMI condition, indicating a greater level of trust in the AV and quicker reaction, than in the eHMI condition. This finding resonates with previous work (e.g., (*Hensch et al.*, 2020)) that has linked HMIs to higher perceived trustworthiness and intuitiveness of AVs. Although there was no significant differences between eHMI and iHMI, the results showed that iHMI condition led to slightly higher trust and awareness in AVs. The superiority of iHMI could be attributed to its location within the CV driver's primary field of view, which aligns with Endsley's (1988) principles of designing for SA enhancement. By present-
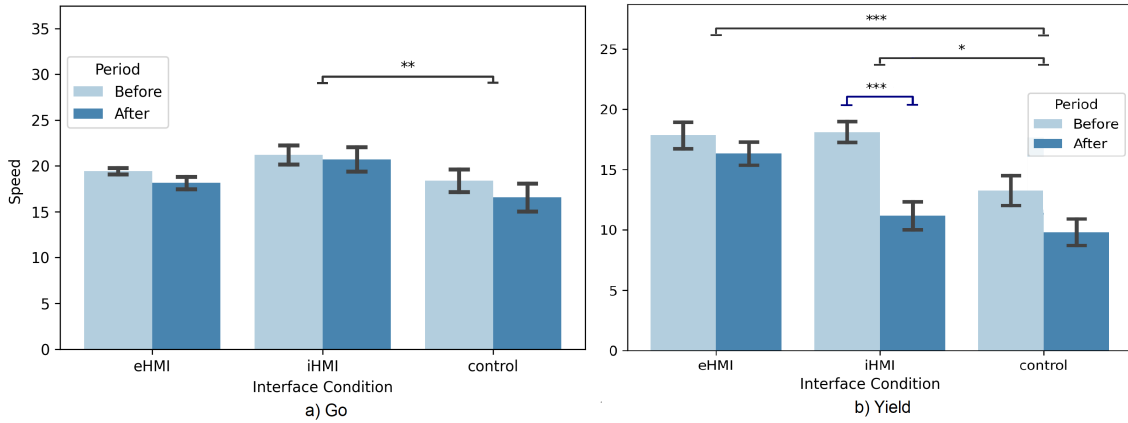
Figure 6.13: Mean speed and standard error across different interface conditions measured in three second time period before and after onset of "Yield" message.

ing information directly on the windshield, iHMIs minimize the need for drivers to divert their attention from the road. In contrast, eHMIs require drivers to shift their focus, which could lead to momentary distractions. This finding suggests that the location and modality of communication interfaces play an important role in their effectiveness, a point that has been underexplored in the AV-CV communication literature.

The HMIs' effectiveness was further supported by our eye-tracking data, which showed a negative increase in mental workload in iHMI condition(see Fig. 6.10) compared to the eHMI and control conditions. This finding aligns with the principles of cognitive load theory (*Sweller*, 2011), which suggest that integrating information into the driver's natural visual field not only improves SA but also reduces the mental effort required to process and act on the information, thereby freeing up mental resources for the primary driving task. This is particularly important in mixed-traffic environments where CV drivers must continuously manage their own vehicle, interpret the intentions of AVs and other road users, and navigate following traffic rules.

With regard to acceptance, both the eHMI and iHMI conditions exhibited high levels of acceptance, surpassing the control condition. However, the iHMI condition was more preferable option, potentially due to the perceived ease of understanding

the shared messages on the interface. Firstly, participants indicated that the iHMI effectively captured their attention immediately, and the messages were easily visible. In contrast, the eHMI required more attentiveness to notice the AV's attempt to communicate, and drivers had to apply additional effort to observe the interface from a distance. This ease of perception with the iHMI translated into faster response times, as evidenced by the more pronounced changes in vehicle speed when participants received the "yield" message (see Fig.6.12). Secondly, the iHMI condition provided clarity regarding the intended recipient of the message. Participants recognized that the displayed actions (i.e., yield or go) were conveyed from the CV driver's perspective, unlike the eHMI condition where the intended recipient could be misinterpreted as other road users. This finding is consistent with participants' ratings and aligns with previous studies (*Eisma et al.*, 2021) on interface effectiveness, confirming that the directness and message perspective in communication fostered a more intuitive and responsive experience for drivers.

Overall, our findings offer valuable insights for the design of AV systems and the management of future mixed-traffic environments. While V2X communication and eHMIs have been the primary focus for enhancing AV-CV communication, our study demonstrates that iHMIs could enhance SA, reduce cognitive load, and increase trust and acceptance. , and supports the Schieben et al.'s (2019) call for designing AV communications based on human needs and expectations.

6.4.1 Implications

Effective communication between vehicles is a vital aspect of developing collaborative driving in a mixed traffic environment. However, due to the complexity of understanding the capabilities and intentions of AVs, it has become more challenging for drivers in CVs to maintain the necessary level of SA, which is critical to ensure transportation safety in mixed traffic. This study showed that both external and

internal HMIs can significantly improve SA during AV-CV interactions at intersections. Sharing the traffic situation and the AV's intention via appropriate designed HMIs boosted human-drivers' SA. Our findings suggest that internal HMIs (iHMIs) may offer advantages over external HMIs (eHMIs). While the iHMI condition is not currently implementable in existing vehicles, it serves as a valuable tool for understanding drivers' cognitive and behavioral responses to AV-provided information. The effectiveness of heads-up display based information visualization, in particular, indicates a promising direction for future research and development. However, the lack of standardized communication protocols remains a significant challenge, potentially leading to confusion and misinterpretation of signals. The ambiguity in AV-CV communication can raise concerns among road users about the reliability and safety of autonomous systems, highlighting the importance of securing public approval and confidence as AV technology advances. Encouragingly, our proposed HMI designs increased trust in AVs compared to the control condition. Nevertheless, it is important to consider that extra information in ambiguous scenarios could potentially add workload to driving tasks, emphasizing the need for balanced design in vehicle communication systems to support real-time decision-making. Our human-centered design approach provides a foundation for addressing these issues, but further research is needed to explore the potential of iHMIs and develop standardized communication protocols. This should encompass various design elements such as visual cues, auditory signals, and optimal HMI placement. Ultimately, the practical implementation of such systems will require collaboration among stakeholders (e.g., automotive manufacturers, technology developers, and policymakers) to establish protocols and ensure equitable access, contributing to the safe integration of AVs into existing transportation infrastructure.

6.4.2 Limitations

This research also has imitations that can be addressed in the future studies. While we proposed several options for HMI designs, the visual accessibility issues (e.g. color-blindness) was not considered. In future studies, the effectiveness of selected interface needs to be validated and refined to accommodate a wider range of users, including those with various visual impairments. As for the second part, first, more objective measures(e.g., eye fixation on areas on interests) should be collected to better understand attention requirements of each design. Second, only one particular scenario was investigated to understand the effects of the proposed HMIs. In future studies, more scenarios should be included to generalize the results for various ambiguous traffic situations. Third, the study was limited by its focus on instruction-based communication between a single AV and CV, which may not fully represent complex real-world traffic scenarios. To address these limitations, future work should explore alternative communication strategies, such as conveying AV intent in environments with multiple road users including CVs and pedestrians. Additionally, we plan to investigate the effectiveness of these strategies in various traffic conditions and urban settings to develop more robust and versatile AV-to-human communication protocols. Forth, the study population mainly included students and was not gender balanced. Future studies should include a more diverse sample to better understanding the effects of HMIs on CV-AV communications in mixed traffic.

6.5 Conclusions

In this research, we aimed to develop an HMI to communicate AV's intention with CV drivers, and investigate how such interfaces would influence CV drivers' SA level and boost their comfort and trust in a mixed traffic environment at intersections. We designed eight different interface concepts and tested the highest-rated concept in two interface conditions (i.e., internal and external HMI) with a control condition. The

effectiveness of the HMIs was evaluated using SA, trust, acceptance, and cognitive workload using participants' eye-tracking measures in ambiguous situations where the CV needed to make a left turn at an intersection with malfunctioning traffic lights. We found that HMIs were assisting CV drivers in uncertain situations and resulted in increase of SA level and trust. The iHMI was considered the most effective communication method with AV and resulted in lowest change in drivers' mental workload.

Chapter 7 Conclusion

7.1 Summary

The proposed research aims to address the communication problem of AVs with other road users in mixed traffic by designing and evaluating explanations that will: 1) establish multdirectional SA dynamically in real time between the AV and road users and 2) improve the SA of both AV drivers and CV drivers in mixed traffic. To accomplish the research objectives, the dissertation was divided into four parts.

First, we investigated the human information needs required to fulfill the cognitive requirements for understanding the environment, as well as the transparency requirements of AVs to comprehend the behavior of intelligent systems. Drawing on theoretical foundations, we constructed a three-level structure that formulates the content of the explanations, addressing the "what" aspect. Subsequently, we defined principles on "how" to deliver the explanations effectively, ensuring prompt and accurate responses. Lastly, we incorporated a dynamic SA assessment component into the framework, enabling objective tracking of SA changes and identifying the optimal timing for providing explanations. All in all, we developed a comprehensive framework that facilitates SA to improve communication among road users in mixed traffic.

Second, we developed a computational model using ensemble and deep learning algorithms to predict the SA in real-time. The study findings helped to understand

the main characteristics describing driver's SA, as well as provided with a method to track the dynamic changes in SA which can help to determine the threshold when the explanation will be necessary.

Third, we conducted a study to examine explanations in AV-driver communication context during conditionally automated driving, utilizing the aforementioned SA framework. The findings demonstrated that the proposed framework was able to alter participants' understanding of traffic situations and enhance their comprehension of AVs. The increased SA contributed to a better understanding of the AV's behavior in specific scenarios, thereby influencing participants' trust in and acceptance of automated vehicles. Furthermore, the results revealed that cognitive requirements varied based on the amount of information presented.

Fourth, we validated the framework's efficiency in an inter-vehicle communication context. To achieve this, We conducted a VR study to enhance SA for CV drivers in mixed traffic environment using HMIs developed by framework and human-centered design principles. We proposed both external and internal HMI concepts to explore the interaction between AVs and CVs in challenging situations, aiming to improve the SA of CV drivers. The results, obtained through self-reported and physiological measures, indicated that explanations successfully increased participants' SA and trust in AVs. Additionally, among the tested communication interfaces, the internal HMI was perceived as the most effective and resulted in minimal changes in drivers' mental workload.

7.2   Contribution

In response to the critical need for effective communication between AVs and other road users in mixed traffic, this work focused on pioneering solutions to improve SA for both AV and other road users in real-time.

This research contributes to our understanding of how to develop multidirectional

SA for enhancing human-machine performance. By designing and evaluating human-centered explanations tailored to individuals' mental models, particularly in conditionally automated driving scenarios, it paves the way for fostering a more harmonious relationship between humans and technology in automated driving and other domains, such as manufacturing and medical industries.

Furthermore, the research holds the potential to enhance road safety and traffic efficiency by facilitating effective communication between autonomous and conventional vehicles in mixed traffic environments. It can accelerate the adoption of self-driving technologies by increasing public understanding and trust in AVs. The development of user-friendly HMIs can significantly improve the usability of in-car systems. Moreover, the proposed computational model has the potential to revolutionize real-time monitoring and control systems in AV, providing a sophisticated framework for seamless integration of autonomous and human-driven vehicles.

This research provides valuable insights into developing and validating a robust communication framework. The strategies employed for understanding human information needs and delivering effective explanations can serve as a guiding principle for leveraging theory-of-mind networks and explainable AI in AVs. Essentially, the project lays the groundwork for building a comprehensive communication framework that can bridge the gap between humans and intelligent systems.

7.3 Future work

Although the conducted investigation helped in enhancing the communication between AV and other road users in mixed traffic, there are several limitation that should be considered in the future. Firstly, to ensure the robustness of the findings, it is recommended to replicate the conducted studies in high-fidelity or naturalistic driving setups. This would provide a more realistic and comprehensive evaluation of the proposed framework.

Second, future studies should include additional physiological and behavioral measures to objectively assess the effectiveness of the proposed approach.

Third, in order to broaden the applicability of the findings, it is important to test the framework in various scenarios encompassing different levels of risk and complexity. This would allow for a more comprehensive assessment of the framework's performance across a diverse range of driving situations.

Finally, future studies should aim to include more diverse sample groups to enhance the generalizability of the results. By incorporating participants from a wider range of backgrounds and demographics, researchers can obtain a more representative understanding of the impact of the proposed framework.

Bibliography

Bibliography

Aoki, S., and R. Rajkumar (2018), Dynamic intersections and self-driving vehicles, in *2018 ACM/IEEE 9th International Conference on Cyber-Physical Systems (ICCPS)*, pp. 320–330, IEEE.

Avetisyan, L., J. Ayoub, and F. Zhou (2022), Investigating explanations in conditional and highly automated driving: the effects of situation awareness and modality, *Transportation research part F: traffic psychology and behaviour*, *89*, 456–466.

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

Ayoub, J., F. Zhou, S. Bao, and X. J. Yang (2019a), 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.

Ayoub, J., F. Zhou, S. Bao, and X. J. Yang (2019b), 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.

Ayoub, J., X. J. Yang, and F. Zhou (2021), Modeling dispositional and initial learned trust in automated vehicles with predictability and explainability, *Transportation Research Part F: Traffic Psychology and Behaviour*, *77*, 102–116.

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

Bai, S., D. D. Legge, A. Young, S. Bao, and F. Zhou (2021), Investigating external interaction modality and design between automated vehicles and pedestrians at crossings, in *2021 IEEE International Intelligent Transportation Systems Conference (ITSC)*, pp. 1691–1696, IEEE.

Bhavsar, P., P. Das, M. Paugh, K. Dey, and M. Chowdhury (2017), Risk analysis of autonomous vehicles in mixed traffic streams, *Transportation Research Record*, *2625*(1), 51–61.

Brown, B., M. Broth, and E. Vinkhuyzen (2023), The halting problem: Video analysis of self-driving cars in traffic, in *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*, pp. 1–14.

Cannon-Bowers, J. A., E. Salas, and S. Converse (1993), Shared mental models in expert team decision making.

Chen, J. Y., K. Procci, M. Boyce, J. L. Wright, A. Garcia, and M. Barnes (2014), Situation awareness-based agent transparency, *Tech. rep.*, ARMY RESEARCH LAB ABERDEEN PROVING GROUND MDUNIVERSITY OF CENTRAL FLORIDA . . . .

Chen, Q., X. Ma, S. Tang, J. Guo, Q. Yang, and S. Fu (2019), F-cooper: Feature based cooperative perception for autonomous vehicle edge computing system using 3d point clouds, in *Proceedings of the 4th ACM/IEEE Symposium on Edge Computing*, pp. 88–100.

Choi, J. K., and Y. G. Ji (2015), Investigating the importance of trust on adopting an autonomous vehicle, *International Journal of Human-Computer Interaction*, *31*(10), 692–702.

Clark, H., A. C. McLaughlin, and J. Feng (2017), Situational awareness and time to takeover: Exploring an alternative method to measure engagement with high-level automation, *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *61*(1), 1452–1456, doi:10.1177/1541931213601848.

Colley, M., and E. Rukzio (2020), A design space for external communication of autonomous vehicles, in *12th International Conference on Automotive User Interfaces and Interactive Vehicular Applications*, pp. 212–222.

De Winter, J. C., R. Happee, M. H. Martens, and N. A. Stanton (2014), Effects of adaptive cruise control and highly automated driving on workload and situation awareness: A review of the empirical evidence, *Transportation research part F: traffic psychology and behaviour*, *27*, 196–217.

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

Dietrich, A., P. Maruhn, L. Schwarze, and K. Bengler (2019a), Implicit communication of automated vehicles in urban scenarios: Effects of pitch and deceleration on pedestrian crossing behavior, in *International Conference on Human Systems Engineering and Design: Future Trends and Applications*, pp. 176–181, Springer.

Dietrich, A., J.-H. Willrodt, K. Wagner, and K. Bengler (2019b), Projection-based external human machine interfaces-enabling interaction between automated vehicles and pedestrians, in *Proceedings of the DSC 2018 Europe VR*.

Du, N., J. Ayoub, F. Zhou, A. Pradhan, L. Robert Jr, D. Tilbury, E. Pulver, and X. J. Yang (2019a), Examining the impacts of drivers' emotions on takeover readiness and performance in highly automated driving, *Proceedings of the Human Factors and Ergonomics Society Annual Meeting.*

Du, N., J. Haspiel, Q. Zhang, D. Tilbury, A. K. Pradhan, X. J. Yang, and L. P. Robert Jr (2019b), Look who's talking now: Implications of av's explanations on driver's trust, av preference, anxiety and mental workload, *Transportation research part C: emerging technologies*, *104*, 428–442.

Du, N., X. J. Yang, and F. Zhou (2020a), Psychophysiological responses to takeover requests in conditionally automated driving, *Accident Analysis & Prevention*, *148*, 105,804.

Du, N., F. Zhou, E. M. Pulver, D. M. Tilbury, L. P. Robert, A. K. Pradhan, and X. J. Yang (2020b), 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., F. Zhou, E. M. Pulver, D. M. Tilbury, L. P. Robert, A. K. Pradhan, and X. J. Yang (2020c), Predicting driver takeover performance in conditionally automated driving, *Accident Analysis & Prevention*, *148*, 105,748.

Du, N., F. Zhou, D. Tilbury, P. L. Robert, and X. J. Yang (2021), Designing alert systems in takeover transitions: The effects of display information and modality, in *Proceedings of the 13th international conference on automotive user interfaces and interactive vehicular applications*, pp. 1–13.

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

Ehrhardt, S., N. Merat, M. Daly, A. S. Crusat, and B. Deml (2024), Comparing implicit communication via longitudinal driving dynamics: a cross-cultural study in germany and the uk, *Transportation research part F: traffic psychology and behaviour*, *102*, 278–293.

Eisma, Y., A. Reiff, L. Kooijman, D. Dodou, and J. De Winter (2021), External human-machine interfaces: Effects of message perspective, *Transportation research part F: traffic psychology and behaviour*, *78*, 30–41.

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, SAGE Publications Sage CA: Los Angeles, CA.

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

Endsley, M. R. (1995b), Measurement of situation awareness in dynamic systems, *Human factors*, *37*(1), 65–84.

Endsley, M. R. (2018), Automation and situation awareness, in *Automation and human performance: Theory and applications*, pp. 163–181, CRC Press.

Endsley, M. R. (2019), Situation awareness in future autonomous vehicles: Beware of the unexpected, in *Proceedings of the 20th Congress of the International Ergonomics Association (IEA 2018)*, edited by S. Bagnara, R. Tartaglia, S. Albolino, T. Alexander, and Y. Fujita, pp. 303–309, Springer International Publishing, Cham.

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

Fernandez Rojas, R., E. Debie, J. Fidock, M. Barlow, K. Kasmarik, S. Anavatti, M. Garratt, and H. Abbass (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, Springer.

Frison, Anna-Katharina, Wintersberger, Philipp, Liu, Tianjia, Riener, and Andreas (2019a), Why do you like to drive automated? a context-dependent analysis of highly automated driving to elaborate requirements for intelligent user interfaces, in *Proceedings of the 24th International Conference on Intelligent User Interfaces*, IUI '19, p. 528–537, Association for Computing Machinery, New York, NY, USA.

Frison, A.-K., P. Wintersberger, T. Liu, and A. Riener (2019b), Why do you like to drive automated? a context-dependent analysis of highly automated driving to elaborate requirements for intelligent user interfaces, in *Proceedings of the 24th international conference on intelligent user interfaces*, pp. 528–537.

Fuest, T., L. Michalowski, L. Träris, H. Bellem, and K. Bengler (2018), Using the driving behavior of an automated vehicle to communicate intentions-a wizard of oz study, in *2018 21st international conference on intelligent transportation systems (ITSC)*, pp. 3596–3601, IEEE.

Garcia, R. (2018), Video shows uber operator moments before self-driving car crash that killed pedestrian, `https://www.usatoday.com/story/tech/nation-now/2018/03/21/fatal-uber-crash/447770002`.

Guanetti, J., Y. Kim, and F. Borrelli (2018), Control of connected and automated vehicles: State of the art and future challenges, *Annual reviews in control*, *45*, 18–40.

Ha, T., S. Kim, D. Seo, and S. Lee (2020), Effects of explanation types and perceived risk on trust in autonomous vehicles, *Transportation Research Part F: Traffic Psychology and Behaviour*, *73*, 271–280.

Haspiel, J., N. Du, J. Meyerson, L. P. Robert Jr, D. Tilbury, X. J. Yang, and A. K. Pradhan (2018), Explanations and expectations: Trust building in automated vehicles, in *Companion of the 2018 ACM/IEEE International Conference on Human-Robot Interaction*, pp. 119–120.

Hensch, A.-C., I. Neumann, M. Beggiato, J. Halama, and J. F. Krems (2020), How should automated vehicles communicate?–effects of a light-based communication approach in a wizard-of-oz study, in *Advances in Human Factors of Transportation: Proceedings of the AHFE 2019 International Conference on Human Factors in Transportation, July 24-28, 2019, Washington DC, USA 10*, pp. 79–91, Springer.

Hoffman, R., S. T. Mueller, G. Klein, and J. Litman (2018), Metrics for explainable ai: Challenges and prospects, *ArXiv, abs/1812.04608*.

Holthausen, B. E., P. Wintersberger, B. N. Walker, and A. Riener (2020a), 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.

Holthausen, B. E., P. Wintersberger, B. N. Walker, and A. Riener (2020b), 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., H. Xie, and E. R. Galea (2018), Perceptions of autonomous vehicles: Relationships with road users, risk, gender and age, *Safety science, 102*, 1–13.

IDEO (2015), *The field guide to human-centered design: Design kit.*, IDEO, San Francisco, CA, U.S.

Isele, D., R. Rahimi, A. Cosgun, K. Subramanian, and K. Fujimura (2018), Navigating occluded intersections with autonomous vehicles using deep reinforcement learning, in *2018 IEEE international conference on robotics and automation (ICRA)*, pp. 2034–2039, IEEE.

Jayaraman, S. K., C. Creech, D. M. Tilbury, X. J. Yang, A. K. Pradhan, K. M. Tsui, and L. P. Robert Jr (2019), Pedestrian trust in automated vehicles: Role of traffic signal and av driving behavior, *Frontiers in Robotics and AI, 6*, 117.

Jensen, T., M. M. H. Khan, Y. Albayram, M. A. A. Fahim, R. Buck, and E. Coman (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.

Juhlin, O. (1999), Traffic behaviour as social interaction-implications for the design of artificial drivers, in *PROCEEDINGS OF 6TH WORLD CONGRESS ON INTELLIGENT TRANSPORT SYSTEMS (ITS), HELD TORONTO, CANADA, NOVEMBER 8-12, 1999*.

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

Kennedy, R. S., N. E. Lane, K. S. Berbaum, and M. G. Lilienthal (1993), Simulator sickness questionnaire: An enhanced method for quantifying simulator sickness, *The international journal of aviation psychology*, *3*(3), 203–220.

Kenney, J. B. (2011), Dedicated short-range communications (dsrc) standards in the united states, *Proceedings of the IEEE*, *99*(7), 1162–1182.

Khastgir, S., S. Birrell, G. Dhadyalla, H. Sivencrona, and P. Jennings (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., N. T. Banerjee, J. Y. Zhang, A. P. Anderson, and T. K. Clark (2023), Estimation of subjectively reported trust, mental workload, and situation awareness using unobtrusive measures, *Human Factors*, *65*(6), 1142–1160.

Koo, J. Kwac, W. Ju, M. Steinert, L. Leifer, and C. Nass (2015), Why did my car just do that? explaining semi-autonomous driving actions to improve driver understanding, trust, and performance, *International Journal on Interactive Design and Manufacturing (IJIDeM)*, *9*, 269–275.

Koo, Jeamin, Shin, Dongjun, Steinert, Martin, Leifer, and Larry (2016), Understanding driver responses to voice alerts of autonomous car operations, *International journal of vehicle design*, *70*(4), 377–392.

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

Körber, M., L. Prasch, and K. Bengler (2018b), Why do i have to drive now? post hoc explanations of takeover requests, *Human factors*, *60*(3), 305–323.

Kuhn, H. W., and A. W. Tucker (1953), *Contributions to the Theory of Games*, 28, Princeton University Press.

Lee, J., T. Daimon, and S. Kitazaki (2021), Negative effect of external human-machine interfaces in automated vehicles on pedestrian crossing behaviour: A virtual reality experiment, in *Congress of the International Ergonomics Association*, pp. 718–725, Springer.

Li, M., B. E. Holthausen, R. E. Stuck, and B. N. Walker (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., P. Blythe, W. Guo, and A. Namdeo (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., L. Zhao, K. Gao, Y. An, and J. Andric (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., J. Yang, K. O. Prakah-Asante, R. Curry, M. Blommer, R. Swaminathan, D. Yu, and B. J. Pitts (2021a), 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, SAGE Publications Sage CA: Los Angeles, CA.

Liang, N., J. Yang, D. Yu, K. O. Prakah-Asante, R. Curry, M. Blommer, R. Swaminathan, and B. J. Pitts (2021b), Using eye-tracking to investigate the effects of pre-takeover visual engagement on situation awareness during automated driving, *Accident Analysis & Prevention*, *157*, 106,143.

Lundberg, S. M., and S.-I. Lee (2017a), A unified approach to interpreting model predictions, in *Advances in Neural Information Processing Systems 30*, edited by I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett, pp. 4765–4774, Curran Associates, Inc.

Lundberg, S. M., and S.-I. Lee (2017b), A unified approach to interpreting model predictions, in *Advances in neural information processing systems*, pp. 4765–4774.

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

Mackay, A., et al. (2019), The impact of autonomous vehicles' active feedback on trust, pp. 342–352, doi:$10.1007/978-3-030-20497-6_32$.

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

Merat, et al. (2019), The "out-of-the-loop" concept in automated driving: proposed definition, measures and implications, *Cognition,Technology and Work*, *21*, doi: 10.1007/s10111-018-0525-8.

Merat, N., T. Louw, R. Madigan, M. Wilbrink, and A. Schieben (2018), What externally presented information do vrus require when interacting with fully automated road transport systems in shared space?, *Accident Analysis & Prevention*, *118*, 244–252.

Miller, L., J. Leitner, J. Kraus, and M. Baumann (2022), Implicit intention communication as a design opportunity for automated vehicles: Understanding drivers' interpretation of vehicle trajectory at narrow passages, *Accident Analysis & Prevention*, *173*, 106,691.

Mohammad, S. H., H. Farah, and A. Zgonnikov (2024), In the driver's mind: modeling the dynamics of human overtaking decisions in interactions with oncoming automated vehicles, *arXiv preprint arXiv:2403.19637*.

Muir, B. M. (1987), Trust between humans and machines, and the design of decision aids, *International journal of man-machine studies*, *27*(5-6), 527–539.

Najm, W. G., J. D. Smith, M. Yanagisawa, et al. (2007), Pre-crash scenario typology for crash avoidance research, *Tech. rep.*, United States. National Highway Traffic Safety Administration.

Naujoks, F., C. Mai, and A. Neukum (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., Z. Zhang, Y. Sun, X. Wang, J. Ni, and H. Qin (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., E. Hu, S. Sheng, S. Kraus, S. Heo, and L. Feng (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.

Palinko, O., and A. L. Kun (2012), Exploring the effects of visual cognitive load and illumination on pupil diameter in driving simulators, in *Proceedings of the Symposium on Eye Tracking Research and Applications*, pp. 413–416.

Palmeiro, A. R., S. van der Kint, L. Vissers, H. Farah, J. C. de Winter, and M. Hagenzieker (2018), Interaction between pedestrians and automated vehicles: A wizard of oz experiment, *Transportation research part F: traffic psychology and behaviour*, *58*, 1005–1020.

Papakostopoulos, V., D. Nathanael, E. Portouli, and A. Amditis (2021), Effect of external hmi for automated vehicles (avs) on drivers' ability to infer the av motion intention: A field experiment, *Transportation research part F: traffic psychology and behaviour*, *82*, 32–42.

Pauzié, A. (2008), A method to assess the driver mental workload: The driving activity load index (dali), *IET Intelligent Transport Systems*, *2*(4), 315–322.

Perello-March, J. R., C. G. Burns, R. Woodman, M. T. Elliott, and S. A. Birrell (2021), Driver state monitoring: Manipulating reliability expectations in simulated

automated driving scenarios, *IEEE transactions on intelligent transportation systems*, *23*(6), 5187–5197.

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

Petersen, Luke, Robert, Lionel, Yang, X. Jessie, Tilbury, and Dawn (2019), Situational awareness, driver's trust in automated driving systems and secondary task performance, *SAE International Journal of Connected and Automated Vehicles*, *2*(12-02-02-0009).

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

Radlmayr, et al. (2014), How traffic situations and non-driving related tasks affect the take-over quality in highly automated driving, in *Proceedings of the human factors and ergonomics society annual meeting*, vol. 58, pp. 2063–2067, Sage Publications Sage CA: Los Angeles, CA.

Rasmussen, J. (1983), Skills, rules, and knowledge; signals, signs, and symbols, and other distinctions in human performance models, *IEEE transactions on systems, man, and cybernetics*, (3), 257–266.

Rasouli, A., and J. K. Tsotsos (2019), Autonomous vehicles that interact with pedestrians: A survey of theory and practice, *IEEE transactions on intelligent transportation systems*, *21*(3), 900–918.

Rettenmaier, M., M. Pietsch, J. Schmidtler, and K. Bengler (2019), Passing through the bottleneck-the potential of external human-machine interfaces, in *2019 IEEE Intelligent Vehicles Symposium (IV)*, pp. 1687–1692, IEEE.

Rettenmaier, M., D. Albers, and K. Bengler (2020), After you?!–use of external human-machine interfaces in road bottleneck scenarios, *Transportation research part F: traffic psychology and behaviour*, *70*, 175–190.

Risto, M., C. Emmenegger, E. Vinkhuyzen, M. Cefkin, and J. Hollan (2017), Human-vehicle interfaces: The power of vehicle movement gestures in human road user coordination, in *Driving assessment conference*, vol. 9, University of Iowa.

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.

SAE, S. (2014), International: on-road automated vehicle standards committee, *Taxonomy and definitions for terms related to on-road motor vehicle automated driving systems, information report*.

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

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

Sanneman, L., and J. A. Shah (2020), A situation awareness-based framework for design and evaluation of explainable ai, in *Explainable, Transparent Autonomous Agents and Multi-Agent Systems*, edited by D. Calvaresi, A. Najjar, M. Winikoff, and K. Främling, pp. 94–110, Springer International Publishing, Cham.

Schewe, F., and M. Vollrath (2020), Ecological interface design effectively reduces cognitive workload–the example of hmis for speed control, *Transportation research part F: traffic psychology and behaviour*, *72*, 155–170.

Schieben, A., M. Wilbrink, C. Kettwich, R. Madigan, T. Louw, and N. Merat (2019), Designing the interaction of automated vehicles with other traffic participants: design considerations based on human needs and expectations, *Cognition, Technology & Work*, *21*(1), 69–85.

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

Sebok, and Wickens (2017), Implementing lumberjacks and black swans into model-based tools to support human–automation interaction, *Human factors*, *59*(2), 189–203.

Seppelt, B. D., and J. D. Lee (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.

Shladover, S. E. (2018), Connected and automated vehicle systems: Introduction and overview, *Journal of Intelligent Transportation Systems*, *22*(3), 190–200.

Siebinga, O., A. Zgonnikov, and D. A. Abbink (2023), Modelling communication-enabled traffic interactions, *Royal Society open science*, *10*(5), 230,537.

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

Smith, K., T. Endsley, and T. Clark (2023), Physiological correlates of objective situation awareness measurements, in *2023 IEEE Aerospace Conference*, pp. 1–6, IEEE.

Society of Automotive Engineers (2018), Taxonomy and Definitions for Terms Related to On-Road Motor Vehicle Automated Driving Systems.

Stanton, N., K. M. Revell, and P. Langdon (2021), *Designing interaction and interfaces for automated vehicles: User-centred ecological design and testing*, CRC Press.

Sweller, J. (2011), Cognitive load theory, in *Psychology of learning and motivation*, vol. 55, pp. 37–76, Elsevier.

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., P. E. Hemeren, and M. Nilsson (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, IEEE.

Tsimhoni, O., P. Green, and J. Lai (2001), Listening to natural and synthesized speech while driving: Effects on user performance, *International Journal of Speech Technology*, *4*(2), 155–169.

Unni, A., K. Ihme, M. Jipp, and J. W. Rieger (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.

van den Beukel, A. P., and M. C. van der Voort (2017), How to assess driver's interaction with partially automated driving systems – a framework for early concept assessment, *Applied Ergonomics*, *59*, 302–312.

van den Beukel, A. P., and M. C. van der Voort (2017), How to assess driver's interaction with partially automated driving systems–a framework for early concept assessment, *Applied ergonomics*, *59*, 302–312.

Van Der Laan, J. D., A. Heino, and D. De Waard (1997), A simple procedure for the assessment of acceptance of advanced transport telematics, *Transportation Research Part C: Emerging Technologies*, *5*(1), 1–10.

Vicente, K. J., and J. Rasmussen (1992), Ecological interface design: Theoretical foundations, *IEEE Transactions on systems, man, and cybernetics*, *22*(4), 589–606.

Vinkhuyzen, E., and M. Cefkin (2016), Developing socially acceptable autonomous vehicles, in *Ethnographic Praxis in Industry Conference Proceedings*, vol. 2016, pp. 522–534, Wiley Online Library.

Wang, Y. L., and F. C. Sus Lundgren, Lyckvi (2020), How drivers respond to visual vs. auditory information in advisory traffic information systems, *Behaviour & Information Technology*, *39*(12), 1308–1319.

Weiner, B. (1979), A theory of motivation for some classroom experiences., *Journal of educational psychology*, *71*(1), 3.

Wickens, C. D. (2008a), Multiple resources and mental workload, *Human Factors*, *50*(3), 449–455.

Wickens, C. D. (2008b), Multiple resources and mental workload, *Human factors*, *50*(3), 449–455.

Wiegand, Gesa, Eiband, Malin, M. Haubelt, Hussmann, and Heinrich (2020), "i'd like an explanation for that!" exploring reactions to unexpected autonomous driving, in *22nd International Conference on Human-Computer Interaction with Mobile Devices and Services*, pp. 1–11.

Wintersberger, Philipp, Frison, Anna-Katharina, A. Riener, Sawitzky, and T. von (2019), Fostering user acceptance and trust in fully automated vehicles: Evaluating the potential of augmented reality, *PRESENCE: Virtual and Augmented Reality*, *27*(1), 46–62.

Yahoodik, S., and Y. Yamani (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., H. Sun, Y. Liu, F. Trapsilawati, O. Sourina, C.-H. Chen, W. Mueller-Wittig, and W. T. Ang (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, IEEE.

Youssef, P., K. L. Plant, and B. Waterson (2024), Narrow passage interactions: a uk-based exploratory survey study to identify factors affecting driver decision-making, *Transportation research part F: traffic psychology and behaviour*, *100*, 402–418.

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

Zgonnikov, A., N. Beckers, A. George, D. Abbink, and C. Jonker (2024), Nudging human drivers via implicit communication by automated vehicles: Empirical evidence and computational cognitive modeling, *International Journal of Human-Computer Studies*, *185*, 103,224.

Zhang, J., E. Vinkhuyzen, and M. Cefkin (2018), Evaluation of an autonomous vehicle external communication system concept: a survey study, in *Advances in Human Aspects of Transportation: Proceedings of the AHFE 2017 International Conference*

on *Human Factors in Transportation, July 17- 21, 2017, The Westin Bonaventure Hotel, Los Angeles, California, USA 8*, pp. 650–661, Springer.

Zhang, T., D. Tao, X. Qu, X. Zhang, R. Lin, and W. Zhang (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.

Zheng, B., X. Jin, Z. Song, Y. Pei, and X. Ma (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, Springer.

Zhou, F., X. Qu, M. G. Helander, and J. R. Jiao (2011), Affect prediction from physiological measures via visual stimuli, *International Journal of Human-Computer Studies*, *69*(12), 801–819.

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

Zhou, F., X. J. Yang, and X. Zhang (2019a), Takeover Transition in Autonomous Vehicles: A YouTube Study, *International Journal of Human–Computer Interaction*, *0*(0), 1–12, doi:10.1080/10447318.2019.1634317.

Zhou, F., X. J. Yang, and X. Zhang (2019b), Takeover Transition in Autonomous Vehicles: A YouTube Study, *International Journal of Human–Computer Interaction*, *0*(0), 1–12, doi:10.1080/10447318.2019.1634317.

Zhou, F., A. Alsaid, M. Blommer, R. Curry, R. Swaminathan, D. Kochhar, W. Talamonti, L. Tijerina, and B. Lei (2020), Driver fatigue transition prediction in highly automated driving using physiological features, *Expert Systems with Applications*, p. 113204.

Zhou, F., A. Alsaid, M. Blommer, R. Curry, R. Swaminathan, D. Kochhar, W. Talamonti, and L. Tijerina (2021a), Predicting driver fatigue in monotonous automated driving with explanation using gpboost and shap, *International Journal of Human–Computer Interaction*, pp. 1–11.

Zhou, F., X. J. Yang, and J. C. de Winter (2021b), Using eye-tracking data to predict situation awareness in real time during takeover transitions in conditionally automated driving, *IEEE Transactions on Intelligent Transportation Systems*.