Designing Explainable In-vehicle Interfaces for Conditionally Automated Driving: A Holistic Examination with Mixed Method Approaches

MANHUA WANG

Dissertation submitted to the faculty of the Virginia Polytechnic Institute and State University in partial fulfillment of the requirements for the degree of

Doctor of Philosophy
In
Industrial and Systems Engineering

Myounghoon Jeon, Chair
Sheila G. Klauer
Rafael N. C. Patrick
Yiqi Zhang

July 29, 2024
Blacksburg, VA

Keywords: Explainable AI, Automated Vehicles, Transparency, In-Vehicle Intelligent Agent, Human-Machine Interface

Designing Explainable In-vehicle Interfaces for Conditionally Automated Driving: A Holistic Examination with Mixed Method Approaches

MANHUA WANG

ABSTRACT

Automated vehicles (AVs) are promising applications of artificial intelligence (AI). While human drivers benefit from AVs, including long-distance support and collision prevention, we do not always understand how AV systems function and make decisions. Consequently, drivers might develop inaccurate mental models and form unrealistic expectations of these systems, leading to unwanted incidents. Although efforts have been made to support drivers' understanding of AVs through in-vehicle visual and auditory interfaces and warnings, these may not be sufficient or effective in addressing user confusion and overtrust in in-vehicle technologies, sometimes even creating negative experiences. To address this challenge, this dissertation conducts a series of studies to explore the possibility of using the in-vehicle intelligent agent (IVIA) in the form of the speech user interface to support drivers, aiming to enhance safety, performance, and satisfaction in conditionally automated vehicles.

First, two expert workshops were conducted to identify design considerations for general IVIAs in the driving context. Next, to better understand the effectiveness of different IVIA designs in conditionally automated driving, a driving simulator study (n=24) was conducted to evaluate four types of IVIA designs varying by embodiment conditions and speech styles. The findings indicated that conversational agents were preferred and yielded better driving performance, while robot agents caused greater visual distraction. Then, contextual inquiries with 10 drivers owning vehicles with advanced driver assistance systems (ADAS) were conducted to identify user needs and the learning process when interacting with in-vehicle technologies, focusing on interface feedback and warnings. Subsequently, through expert interviews with seven experts from AI, social science, and human-computer interaction domains, design considerations were synthesized for improving the explainability of AVs and preventing associated risks. With information gathered from the first four studies, three types of adaptive IVIAs were developed based on human-automation function allocation and investigated in terms of their effectiveness on drivers' response time, driving performance, and subjective evaluations through a driving simulator study (n=39). The findings indicated that although drivers preferred more information provided to them, their response time to road hazards might be degraded when receiving more information, indicating the importance of the balance between safety and satisfaction.

Taken together, this dissertation indicates the potential of adopting IVIAs to enhance the explainability of future AVs. It also provides key design guidelines for developing IVIAs and constructing explanations critical for safer and more satisfying AVs.

Designing Explainable In-vehicle Interfaces for Conditionally Automated Driving: A Holistic Examination with Mixed Method Approaches

MANHUA WANG

GENERAL AUDIENCE ABSTRACT

Automated vehicles (AVs) are an exciting application of artificial intelligence (AI). While these vehicles offer benefits like helping with long-distance driving and preventing accidents, people often do not understand how they work or make decisions. This lack of understanding can lead to unrealistic expectations and potentially dangerous situations. Even though there are visual and sound alerts in these cars to help drivers, they are not always sufficient to prevent confusion and over-reliance on technology, sometimes making the driving experience worse. To address this challenge, this dissertation explores the use of in-vehicle intelligent agents (IVIAs), in the form of speech assistant, to help drivers better understand and interact with AVs, aiming to improve safety, performance, and overall satisfaction in semi-automated vehicles.

First, two expert workshops helped identify key design features for IVIAs. Then, a driving simulator study with 24 participants tested four different designs of IVIAs varying in appearance and how they spoke. The results showed that people preferred conversational agents, which led to better driving behaviors, while robot-like agents caused more visual distractions. Then, through contextual inquiries with 10 drivers who own vehicles with advanced driver assistance systems (ADAS), I identified user needs and how they learn to interact with in-car technologies, focusing on feedback and warnings. Subsequently, I conducted expert interviews with seven professionals from AI, social science, and human-computer interaction fields, which provided further insights into facilitating the explainability of AVs and preventing associated risks. With the information gathered, three types of adaptive IVIAs were developed based on whether the driver was actively in control of the vehicle, or the driving automation system was in control. The effectiveness of these agents was evaluated through drivers' brake and steer response time, driving performance, and user satisfaction through another driving simulator study with 39 participants. The findings indicate that although drivers appreciated more detailed explanations, their response time to road hazards slowed down, highlighting the need to balance safety and satisfaction.

Overall, this research shows the potential of using IVIAs to make AVs easier to understand and safer to use. It also offers important design guidelines for creating these IVIAs and their speech contents to improve the driving experience.

ACKNOWLEDGEMENTS

Looking back on my four-year PhD journey, I have received tremendous support from my advisor, mentors, family, and friends.

To my advisor, Dr. Myounghoon Jeon: You have always had unwavering faith in me and my abilities, even during times when I doubted myself and felt anxious about uncertainties. Thank you for always believing in me. Whenever I got stuck into research ideas, you were always able to guide me the way out. I will never stop being amazed by how knowledgeable you are. You have not only introduced me to the world of research but also shown me what it means to be a great advisor, which I deeply appreciate and aspire to take the legacy.

To my committee members, Dr. Klauer, Dr. Patrick, and Dr. Zhang: Thank you for sharing your wisdom and guiding me through my dissertation. I have gained invaluable knowledge from your classes and our in-person interactions. Your openness and support have been greatly appreciated.

To all the mentors I have encountered through conferences, internships, cold emails, and interviews: It has been a privilege to receive guidance and support from so many pioneers and leading researchers in the field. Your advice has extended beyond research to include career and life skills, for which I am deeply grateful.

To my cohort at Virginia Tech ISE, especially my lab mates in the Mind Music Machine lab: I've had the pleasure of spending four years with an incredible group of students. I am fortunate to have worked with and gotten to know you all, and some of us have developed deep friendships that I will always cherish.

To my family: My mom and dad, who have unconditionally supported me both mentally and financially; and my uncle and aunt in Chapel Hill, who have been like family to me despite not sharing blood ties. To all of my family members who give me unconditional love!

Finally, to my support system that has upheld me through smiles and tears: Jiayuan Dong, Huayu Liang (Kira), Lingyu Li (Chris), Shiyu Deng, Fayu Chong, Shafiqul Islam, Sharon Dan, Megan Fok, Megan Vogt, Brittany Health, Xiang Yang, Jing Zang, and Jacqueline Bruen. Jia, you cannot imagine how lucky I feel to know and be loved by you. I have grown so much with you and am grateful for your bravery, courage, confidence, and kindness, and so on. I can always learn from you no matter how long we have known each other. Thank you for always being by my side and tolerant my annoyance :) Kira and Chris, you have brought so much joy into my life and shared countless food and life tips. I cannot express how much each of you has contributed to my well-being and success over the years. I will miss you all when we start our own adventures after Virginia Tech. Always and Forever!

ATTRIBUTIONS

Workshops mentioned in Chapter 3, Study 1 were conducted at AutoUI 2021, and AutoUI 2022 conferences:

Wang, M., Hock, P., Lee, S. C., Baumann, M., & Jeon, M. (2021). Genie vs. Jarvis: Characteristics and Design Considerations of In-Vehicle Intelligent Agents. *13th International Conference on Automotive User Interfaces and Interactive Vehicular Applications*, 197–199. https://doi.org/10.1145/3473682.3479720

Wang, M., Hock, P., Lee, S. C., Baumann, M., & Jeon, M. (2021). Genie vs. Jarvis: Characteristics and Design Considerations of In-Vehicle Intelligent Agents. *13th International Conference on Automotive User Interfaces and Interactive Vehicular Applications*, 197–199. https://doi.org/10.1145/3473682.3479720

Portions of Study 1 were adapted from an HFES conference proceeding:

Wang, M., Park, S. H., Lee, S. C., Hock, P., & Baumann, M. (2022). Build Your Own Genie and Jarvis: 2nd Workshop on Characteristics and Design Considerations of In-Vehicle Intelligent Agents. *14th International Conference on Automotive User Interfaces and Interactive Vehicular Applications*, 176–178. https://doi.org/10.1145/3544999.3552313

Portions of Study 2 were adapted from an AutoUI conference proceeding:

Wang, M., Lee, S. C., Montavon, G., Qin, J., & Jeon, M. (2022). Conversational Voice Agents are Preferred and Lead to Better Driving Performance in Conditionally Automated Vehicles. *Proceedings of the 14th International Conference on Automotive User Interfaces and Interactive Vehicular Applications*, *1*(1), 86–95. https://doi.org/10.1145/3543174.3546830

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1 INTRODUCTION

Artificial Intelligence (AI) has been seamlessly integrated into human life through a variety of applications, including natural language processing (e.g., virtual assistants), image and video processing (e.g., object and face recognition), automation systems (e.g., automated vehicles), or predictive analysis (e.g., customer behaviors). Automated vehicles (AVs)—one of the most promising AI systems—have achieved significant technological advancements, attracting substantial research and commercial interest. Although some commercially available AVs can control both longitudinal and lateral vehicle movements and respond to road obstacles, their operations are limited to certain conditions. However, end users usually have difficulties understanding these limitations and, thus, develop inaccurate trust towards AVs. While undertrust leads to a lower acceptance rate and unwillingness to purchase, overtrust can lead to fatalities due to system misuse. Therefore, effectively communicating AV capabilities and limitations to users is crucial for enhancing road safety and reducing future transportation system fatalities.

Given that the level of automation varies along a spectrum, it is anticipated that the information requirements to explain AV capabilities and limitations will need to be adapted accordingly. The International Society of Automobile Engineers (SAE International, 2021) defines six discrete and mutually exclusive levels of driving automation in the context of motor vehicles: Level 0-No driving automation, Level 1-Driver Assistance, Level 2-Partial Driving Automation, Level 3-Conditional Driving Automation, Level 4-High Driving Automation, and Level 5-Full Driving Automation (SAE International, 2021). Each level of driving automation defines the respective roles of human drivers that change as the functionality of the driving automation system alters. The roles of Level 1 and 2 driving automation systems are to perform the sustained longitudinal and/or lateral vehicle motion control subtasks of the dynamic driving task (DDT). Drivers are no longer required to maintain these subtasks but are still expected to complete the DDT by maintaining the remaining vehicle control (if any) and performing the object and event detection and response (OEDR) subtask. A driving automation system that reaches Level 3 carries out the entire DDT. However, a human user is expected to serve as a DDT

fallback-ready user to take over the DDT when receiving alerts from system failures or when the driving automation system is approaching its operational design domain (ODD) limits. Finally, a driving automation system can qualify as Level 4 or Level 5 if it is capable of performing the entire DDT and DDT fallback without mandatory user intervention (Level 4) or without user intervention at all (Level 5). Under these conditions, all in-vehicle users will become passengers. **For the purpose of the present dissertation, I specifically focus on addressing issues existing within lower levels of automation, such as Levels 2-3, because the control allocation is perceived as confusing (Novakazi et al., 2021) under this shared-control condition, which can lead to negative consequences.**

Explanations of the system reasoning process are necessary to help drivers form a proper understanding of AV systems and use them wisely (Noble et al., 2019). Explainable AI (XAI) has been advocated and strategized in many decision-making assistance systems to improve transparency and explainability. Nevertheless, XAI has not been adequately implemented in automated driving systems (ADS) to enhance user understanding and improve human-automation collaboration. Atakishiyev (2023) reviewed 38 studies on XAI-based autonomous driving. Only 11 of them (28.9%) aimed to address challenges for users in addition to AV developers (including road users, regulators and insurers, and executive management of automobile companies), among which only five studies focused on the needs of end users (e.g., road users or drivers) (Atakishiyev et al., 2023). Existing research on this topic has explored how to provide explanations of vehicle intentions and decisions on different activities ranging from dynamic driving tasks (Q. Zhang et al., 2021) to tertiary driving tasks (e.g., interaction with infotainment system) (Graefe et al., 2022), most of which were designed based on researchers' intuition of what constitutes good explanations (Miller, 2019). One of the fundamental principles of XAI – *Meaningful* – has been overlooked or not fully considered. *Meaningful* is not about simply providing explanations but requires explanations to be understandable to the system's intended users, tailored to the user needs and level of expertise, and relevant to the context (Phillips et al., 2021). Providing unnecessary information without

considering explanation needs can adversely affect drivers' trust and performance in using ADS (Wiegand et al., 2020).

The overall structure of this dissertation is shown in Figure 2.1-1. This dissertation first explored the possibility of using intelligent agents to support drivers in conditionally automated driving (Study 1 and Study 2). Subsequently, it implemented two qualitative studies to understand the needs and requirements for explanation under various scenarios (Study 3 and Study 4). With both understanding of IVIA design requirements and explanation design requirements and the existing theoretical frameworks (Endsley, 2016; Lyons, 2013), I designed and evaluated the effects of three alternative in-vehicle intelligent agents to support drivers to form appropriate mental models of Level 3 ADS (Study 5). These three types of agents varied based on whether they were able to adapt when the driver responsibilities changed in conditionally automated driving. The proposed intelligent agents were tested in an advanced driving simulation system, which allowed us to simulate futuristic ADS technologies with higher flexibility and controllability. Drivers' driving behaviors and subjective evaluations were collected to examine the effectiveness of the proposed intelligent agents.

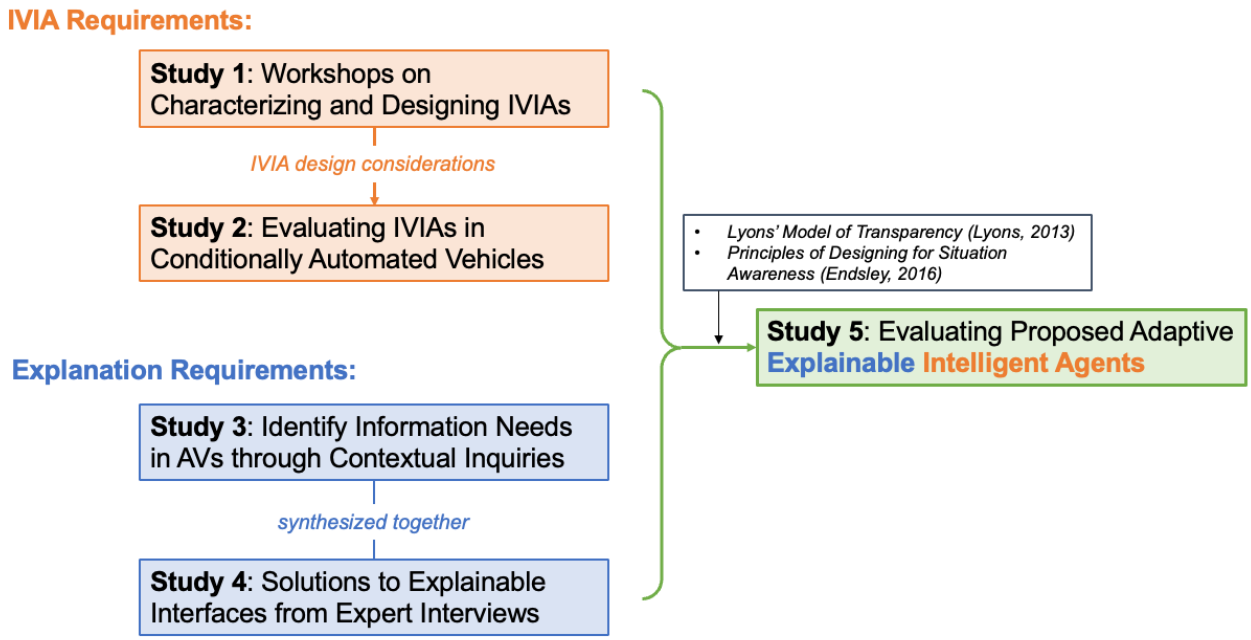

Figure 2.1-1 Dissertation structure: Summary of studies.

The contributions of this dissertation research are four-fold. First, the findings from this dissertation support the importance of using qualitative methods to understand user needs in AVs, which encourages future research to emphasize the information-gathering process. Second, the design guidelines resulting from the qualitative research activities provide future directions for designing explainable interfaces for AVs. Third, the proposed intelligent agents could adapt based on driver-automation function allocation, which fosters the discussion and further investigation of the optimal form factors of adaptive explainable interfaces in AVs. Finally, the findings from this dissertation work contribute to the theoretical framework that advances the fields of human-automation interaction.

In addition to this series of significant findings, the outcomes of this dissertation have the following broader impacts. First, the research outcomes can be used to promote education on human-AI teaming in the context of AVs. Second, findings from this dissertation research can be expanded to other fields, such as healthcare engineering. Finally, the explainable interfaces advocated in the present study are beneficial to vulnerable populations like people with visual impairments.

CHAPTER 2 BACKGROUND AND RELATED WORKS

2.1 Human-Machine Interfaces for Automated Vehicles

To date, the in-vehicle human-machine interfaces (HMIs) are highly versatile. HMIs for automated vehicles (AVs) can be largely divided into two categories: supporting non-driving-related tasks and supporting driving-related tasks. **Supporting non-driving related tasks** includes assistance in manipulating in-vehicle infotainment systems that involve entertainment (e.g., manipulating a radio) and other activities that might pose distractions. According to Geiser (1985)'s classification of driving tasks, HMIs that **support driving-related tasks** can split into supporting primary driving tasks (e.g., navigating, operating, and stabilizing the vehicle), secondary driving tasks (e.g., operating indicators, controlling wipers), and tertiary driving tasks (e.g., adjusting the air conditioners) (Geiser, 1985). Based on the information conveyed (Capallera et al., 2023) that results from the HMI primary functionality, HMIs **supporting primary driving tasks** can be further divided into three categories: **directly assisting driving tasks**, **indicating AV function status**, and **monitoring and intervening driver status**. These HMIs are usually designed to reduce drivers' workload and improve driving safety.

To **assist non-driving-related tasks (NDRTs)**, in-vehicle HMIs allow drivers to manipulate infotainment systems. HMIs in AVs are intended to enable NDRTs beyond classic ones that are already prevalent in manual driving (e.g., making phone calls and texting). More or less, those tasks are typically visual-demanding and should not be performed under manual driving. It is promising that with the assistance of HMIs, drivers are able to perform atypical NDRTs such as video conferencing, gaming, reading and writing, and so on (Bengler et al., 2020).

HMIs designed to **directly assist driving tasks** typically provide traffic-related information and driving maneuver recommendations. The traffic-related information includes surrounding road users (Hartwich et al., 2021; Lau et al., 2020) and traffic control devices (Hartwich et al., 2021). Simply providing HMIs that communicate traffic-related information can improve driver trust and enhance their driving pleasantness

(e.g., perceived safety, system understanding, driving comfort, and driving enjoyment) in AVs, especially under complex situations (Hartwich et al., 2021). Some advanced driving automation systems also provide information related to driving maneuver recommendations, including takeover requests and specific vehicle controls (e.g., brake). With lower levels of automation (Levels 1 and 2) where drivers are responsible for partial dynamic driving tasks (DDTs), adding HMIs that are able to provide recommended responses to road events can significantly improve driving performance (e.g., Hester et al., 2017; Mahajan et al., 2021a). A higher level of automation system (Level 3) is still not able to handle all the road events due to technology and decision-making limitations. Thus, an authority transition from automation to a human driver—termed, "takeover"—is necessary to help drivers handle those situations. More importantly, HMIs that convey additional information (e.g., spatial audio) related to takeover maneuvers yielded better and safer takeover performance (Sanghavi et al., 2021; Stojmenova et al., 2020).

In addition to supporting driving tasks, in-vehicle HMIs **providing information related to AV function status** are often advocated to improve driving safety and driver trust. The AV function status here refers to whether the automation system is ready and safe to use, whether it is functioning properly, its intention, and the reasoning for the AV actions (e.g., lane change, brake, or the system reaches its limitation). A number of commercially available vehicles with Levels 1-2 driving automation systems have already adopted visual elements to indicate the readiness of the automation feature, but the safe-to-use status is not always well communicated (Monticello, 2023). When the AV function is engaged, malfunction or deactivation can occur due to system error or environmental constraints. Providing information about the causes and characteristics of system malfunction in advance can prevent driver trust from decreasing due to system errors (Kraus et al., 2020). Vehicle intention information is also critical for drivers to maintain situation awareness and practice safe driving behaviors (Lau et al., 2020), which can be challenging for distracted drivers due to the automation system engagement (Dunn et al., 2021). Simply adding the AV's intended pathway (i.e., "what" the vehicle is doing) can improve driving safety in AVs because drivers are more capable of intervening with silent system failures, which also boosts driver trust (Swain et al.,

2023). However, only conveying “what” action the vehicle is taking or will take is not sufficient; additional explanation of “why” the AV acts in a certain way is also needed. Solely providing such “why” explanations can promote driver trust and safe driving performance while reducing anxiety (Koo et al., 2015).

Some advanced vehicles are also equipped with technologies that can **monitor driver status and provide timely interventions**. Existing research studies have developed and tested driver monitoring systems through face recognition and eye-tracking using computer vision and deep learning techniques (J. Zhang et al., 2021). In addition, emotion recognition and intervention have also been researched to mitigate the negative effects of emotional driving. After identifying inattentive drivers, the HMIs also perform interventions to bring drivers back to the driving tasks if necessary. The presence of driver monitoring and intervention systems can increase driving safety. For example, social interactions provided by in-vehicle agents using both emotion regulation prompts and situation awareness prompts can reduce anger levels and perceived workload in angry drivers, enhancing driver situation awareness and driving performance (Jeon et al., 2015). Choe and Jeon (2023) compared intelligent agents adopting affective empathy and cognitive empathy strategies on their effectiveness in mitigating anger effects. Their preliminary results indicated that the affective empathy strategy was more effective in mitigating anger and encouraging safe driving behaviors compared to the cognitive empathy strategy (Choe & Jeon, 2023).

From a holistic perspective, all these different types of HMI together contribute to facilitating driver-automation interaction, improving drivers’ trust and task performance. The HMI equipped in the AV can have multiple functions mentioned above (Mehrotra et al., 2022), addressing traffic safety from multiple aspects. However, even with these HMIs equipped, the quality of driver-automation interaction also depends on multiple factors.

The focus of this dissertation is on the HMIs supporting driving-related tasks. Thus, the related work synthesized in the following sections mainly considers elements related to driving-related tasks.

2.2 Factors Influencing Driver-Automation Interaction

Factors that influence driver-automation interaction can be categorized into driver characteristics, technology features, and environmental factors.

2.2.1 Driver characteristics

The main driver characteristics that have been shown to influence driver-automation interaction include driver demographics, such as **age**, **driving experiences**, **digital literacy**, and driver status, such as **attentiveness** and **emotional status**.

In highly automated vehicles, older drivers perform differently compared to younger drivers when using HMIs that were designed to assist the takeover process. Older drivers had the longest takeover time when using the basic HMIs that informed the takeover request without any further information, while the younger drivers had the shortest takeover time when using basic HMIs (Li et al., 2019). The takeover time for older drivers could be reduced by up to 1.3 seconds through the provision of additional information (e.g., vehicle automation system status, reasoning for vehicle decision) (Li et al., 2019), suggesting that older drivers may benefit from extensive information displayed on HMIs. A review of HMI designs for older drivers has emphasized the need for ongoing research to develop HMI design guidelines to effectively accommodate the sensory, cognitive, and physical capabilities and limitations of the older population (Young et al., 2017). This dissertation research focuses on establishing the preliminary guidelines for drivers riding in automated vehicles. Considering that older drivers might require special needs and adjustment for explainable interfaces, I primarily focus on young drivers given the convenience sampling recruited through university student population. The needs of older drivers will be explored in the future.

Little research has been conducted to explore how digital literacy influences drivers' interaction with HMIs. Alongside digital literacy, the concept of AI literacy has also been introduced, recognizing the growing prevalence of AI technologies. The influence of digital literary or AI literacy on HMIs mainly falls under user acceptance and user understanding. Through the lens of the Technology Acceptance Model (Davis, 1986), lower digital literacy and AI literacy might decrease the perceived ease of use, which can be seen from the current research on training older adults to use advanced technologies. As a consequence, people with lower digital and AI literacy might lead to lower user acceptance. On the contrary, people with higher digital literacy and AI literacy but not at the expert level can overestimate the system usefulness, leading to overtrust and system misuse (Parasuraman & Riley, 1997). More research needs to be done to understand the general AI literacy among the public and how it influences people's technology use.

Finally, in addition to the enduring driver demographic factors, driver status while operating a vehicle also has an impact on their interaction with HMIs. Driver distraction has been extensively researched. Interacting with HMIs can be distracting, especially under manual driving conditions. However, in the case of AVs, drivers are susceptible to a multitude of distractions. Information provided by HMIs can attract driver attention promptly and assist them in regaining situation awareness. Distracted drivers require more comprehensive HMIs that provide information about essential vehicle status, other road users of interest, and the intentions of the AV (Lau et al., 2020).

Drivers under emotional impacts also interact with HMIs in different ways compared to those without emotional impacts, providing an opportunity for positive interventions to mitigate negative emotional impacts. Emotions can impact drivers' attention, judgement, and decision-making (Jeon, 2015), making them a crucial consideration in HMI design. In manual driving, HMIs that incorporate affective components have shown promise in reducing drivers' stress (Williams & Breazeal, 2013), promoting safe driving behaviors, and enhancing driver-HMI interaction (Johnsson et al., 2005; Nass et al., 2005). In addition to basic emotions (e.g., happiness, anger), drivers can easily experience boredom if they are not

actively engaging in the driving tasks when situated in AVs. Negative impacts from boredom can be mitigated by in-vehicle intelligent agents that are capable of showing empathy, which further improves drivers' performance (Samrose et al., 2020). Thus, considering the affective aspects of driving and HMI design is also valuable.

2.2.2 Technology features

The AV technology features also play a vital role in altering drive-automation interaction, especially when considering **communication modality** and **information presentation**.

2.2.2.1 Communication Modality.

The communication modality refers to the primary sensation modalities used in the driving context: visual, auditory, tactile, olfactory, and a combination of two or more modalities. Olfactory displays have been used to communicate emotions such as fear (De Groot et al., 2014), which has limited capability in explaining AV intentions and suggestions. Thus, olfactory displays will not be discussed in the following sections.

The most researched HMI function in the AV context is issuing takeover requests (TORs) that inform drivers to take over the control and drive the vehicle should the system reach its operational limitation. Considering that driving is already a visually demanding task, HMIs that primarily use visual displays to deliver information typically compromise the driving tasks and can introduce additional danger. Thus, researchers have also investigated the effectiveness of auditory displays and tactile displays in automated vehicles. In terms of unimodal displays, a driving simulator study with 101 participants found that auditory and tactile TORs are significantly more effective in alerting drivers engaging in non-driving related tasks (NDRT), indicated by faster reaction time and higher perceived usefulness (Petermeijer et al., 2017). Although the effectiveness between auditory and tactile alerts did not differ significantly, tactile alerts were found to be more attention-capturing (Petermeijer et al., 2016). However, TORs using multimodal modalities generally yielded better performance than unimodal TORs (McDonald et al.,

2019). The effectiveness of multimodalities can vary depending on the elements combined in the display. For instance, the auditory-visual TORs yielded better takeover time, but the tactile-visual TOR interface helped with better post-takeover maneuvers and safer driving behaviors (Gruden et al., 2022).

2.2.2.2 Information Presentation.

The information presentation here refers to the way that the contents are delivered. The information presentation is naturally distinct across different modalities. Thus, it is more meaningful to discuss the effectiveness of different information presentation styles within each communication modality.

Visual Display. There are different types of visual cues incorporated in visual HMIs supporting driving-related tasks, for example, pictorial icons, numbers, and texts. Usually, a combination of two or more types of visual cues constitutes a single visual display. In terms of timing, some visual displays provide information that updates in real-time, while other displays present static information and only alert when there is a critical situation. While visual displays can also be versatile, there are some general guidelines to follow, such as the Gestalt Principles (Bennett & Flach, 2011b), Tufte Design Principles (Tufte, 2013), psychophysical approaches (Bennett & Flach, 2011a), etc.

Auditory Display. There are different types of auditory cues used for auditory displays, such as auditory icon, speech, earcon, and spearcon (Jeon et al., 2022; Richie et al., 2018). The **primary** auditory cues used in the auditory displays for HMIs supporting driving-related tasks are earcons (primarily chimes and music) and speech (Capallera et al., 2023), and can lead to faster reaction time and higher user satisfaction compared to spearcon (Jeon, 2019). An earcon has an arbitrary relationship with the object or action that it is referring to and usually is a short, synthetic, and musical tone (Blattner et al., 1989). Speech is verbal communication that is generated by real human voices or synthesized voices. Table 2.2-1 demonstrates how different information (e.g., perceived urgency, event information) is presented for each type of auditory display. Due to its conciseness and benefits from binaural processing, earcons have been researched to convey hazard location through spatiality (Sanghavi et al., 2021; Stojmenova et al., 2020).

They can also be used to convey event urgency by manipulating base frequency, the number of harmonics, and loudness (Lewis et al., 2018; Sanghavi et al., 2020). But when it comes to providing explanations and indicating empathy, earcons are very limited. On the contrary, speech, due to its versatility, can convey not only the location and urgency information (Choe & Jeon, 2023; Taylor et al., 2023; Wong et al., 2019) but also provide explanations when needed and demonstrate empathy characteristics through speech contents, speech style, and tone (Koo et al., 2015; Samrose et al., 2020; Wang, Lee, et al., 2022; Williams & Breazeal, 2013). So far, the exemplar speech presentations mentioned above are all one-way, meaning that the system provides explanations all at once for a certain event prior to user request and does not take further queries (i.e., "proactive" system). There are also two-way systems that present initial pieces of information and provide more information upon request (i.e., "on-demand"). Taylor et al. (2023) compared these two types of explanation styles under different system reliability conditions and found that the on-demand style was perceived as more natural. They also advocated that the system transparency needs to match with automation system reliability to avoid inappropriate trust level developed (Taylor et al., 2023). It is noted that although speech is able to incorporate a large scale of information, the perception of speech information can exert additional cognitive load on drivers to decode and comprehend the messages (Jeon et al., 2022).

Table 2.2-1 Information presentation across earcons, and speech.

Information	Related Functionality	Information Presentation	
		Earcon	Speech
Hazard Location	• Driving tasks	• Spatiality	• Direct contents
Urgency	• Driving tasks • AV status • Driver monitoring	• Base frequency • No. of harmonics • Pulse rate • Loudness	• Keywords (e.g., danger) • Speech Style
Explanations	• Driving tasks • AV status • Driver monitoring	-	• Direct contents • On-demand vs. proactive
Empathy	• Driver monitoring	-	• Speech style • Direct contents • Voice tone

Tactile Display. Tactile information can be coded in four dimensions: frequency, amplitude, location, and duration (Petermeijer et al., 2016). Vibrations with higher frequency and amplitude are more salient and are more efficient in capturing driver attention (Petermeijer et al., 2016). In addition to providing alerts, manipulating the location dimension can add information about the surrounding hazards and instructions about a certain action. However, decoding instructions solely from the tactile pattern can be confusing, so drivers have to rely on other information (e.g., visual cues) to confirm the action (Telpaz et al., 2015). In addition, compared to only patterns without instructions, meaningful instructional tactile information can lead to longer takeover response time because of the complexity of decoding (Huang & Pitts, 2022). Considering the ambiguity of the effectiveness of tactile displays, the present dissertation does not consider them in the proposed HMIs.

2.2.3 Environmental factors

Finally, the environmental factors are seldom discussed but should not be overlooked. Traffic conditions such as traffic volume can influence the effectiveness of HMIs, drivers' adaptation of HMIs, and their preference towards HMI modalities. For example, traffic density has been found to impact driver-HMI interaction and driver performance. Increasing traffic density prolongs drivers' reaction time towards the takeover request in conditionally automated vehicles and also yielded worse takeover quality indicated by greater maximum lateral accelerations (Gold et al., 2016). Increasing the number of objects of interest (e.g., pedestrians and other vehicles) can compromise drivers' situation awareness (Park et al., 2022), which is defined as "the perception of the elements in the environment within a volume of time and space, the comprehension of their meaning, and the projection of their status in the near future" (Endsley, 1988). Thus, drivers' interaction with HMIs can be limited and disturbing in complex environments. However, such critical situations also provide opportunities for HMIs to help drivers comprehend and respond to traffic complexity in a timely and safe manner.

In addition, the safety criticality of the target road events, which is mostly referred to as urgency, also influences drivers' interaction with HMIs. Typically, multimodal HMIs are designed to deliver warnings

that are highly urgent (e.g., takeover requests and driver alertness warnings). Non-urgent messages are also encouraged to help drivers prepare for the interruption of NDRTs (if any) (Naujoks et al., 2017). Previous research also investigated unplanned and planned TORs. The unplanned TORs typically have shorter time budgets for takeover, meaning the situations are more time critical. On the contrary, planned TORs usually provide longer time budgets (e.g., 15 seconds or longer) and are presented in non-safety critical events such as highway exit (Hong et al., 2022; Tan & Zhang, 2022). Hong et al. (2022) found that for unplanned takeover events, the A-pillar LED light strip combined with earcon was better than combined with speech messages in terms of post-takeover performance, indicated by smaller steering angle deviation. However, for planned takeover events, the A-pillar LED light strip without haptic feedback yielded better takeover performance observed from a smaller maximum acceleration, compared to the light incorporating haptic feedback. In a freeway exiting condition, Tan and Zhang (2022) explored the effect of TOR lead time on drivers' situation awareness in conditionally automated vehicles and recommended a takeover time between 25-30 seconds for not only better situation awareness but also higher subjective evaluations on takeover readiness, workload, and trust. The safety criticality also impacts driver responses to automated vehicle interventions. Dixon et al. (2023) found that in safety-critical scenarios, drivers were willing to give automation more control and respond more positively to the automation system interventions that might "override" driver control.

To summarize, not only HMI designs but also the use contexts have influences on human-HMI interaction. Interacting with HMIs is a dynamic and complex task embedded within the driving task, which is worth continuous exploration and adaptation to the advanced automated vehicle technologies.

2.3 Existing Models for Explainable AI (XAI)

Explainable artificial intelligence (XAI) has been brought up by multiple government agencies and received excessive attention for a long time. In 2015, the Defense Advanced Research Projects Agency (DARPA) formulated the XAI program and then launched it in 2017. DARPA defined XAI as "AI systems that can explain their rationale to a human user, characterize their strengths and weaknesses, and

convey an understanding of how they will behave in the future" (Gunning & Aha, 2019). Researchers involved in the XAI programs have worked on three aspects of XAI development: explainable models, explanation interfaces, and psychological models of explanation. The explainable models explored and evaluated in this program mainly focused on the algorithm aspects. The explanation interfaces focused on enhancing the interpretation of the algorithm decision-making process. The psychological model of explanation applies psychological and human-computer interaction (HCI) theories of explanation. My dissertation research focuses on the aspects of explanation interfaces and psychological models of explanation.

The National Institute of Standards and Technology (NIST) later released four principles of XAI: Explanation, Meaningful, Explanation Accuracy, and Knowledge Limits (Phillips et al., 2021). The definitions of each of these principles are listed below:

> ***"Explanation****: A system delivers or contains accompanying evidence or reason(s) for outputs and/or processes.*
>
> ***Meaningful****: A system provides explanations that are understandable to the intended consumer(s).*
>
> ***Explanation Accuracy****: An explanation correctly reflects the reason for generating the output and/or accurately reflects the system's process.*
>
> ***Knowledge Limits****: A system only operates under conditions for which it was designed and when it reaches sufficient confidence in its output."*

In the field of automated vehicles, usually, the automation systems are well restricted to their knowledge limits, meaning that they are typically designed to be operated within their operation domain. In general, automation systems also provide explanations to drivers when they are active. However, whether those pieces of information are meaningful and accurate is questionable—which is the primary issue that motivates this present dissertation research.

2.3.1 Terminology: Transparency, Interpretability, and Explainability

Transparency, interpretability, and explainability have been used in the field of XAI interchangeably, yet these terms are different (Barredo Arrieta et al., 2020; Phillips et al., 2021; Roscher et al., 2020). To avoid confusion, here I distinguish these three terms as follows. An overarching distinction across these terms is that transparency considers the AI algorithm or the Machine Learning (ML) approach, interpretability considers the ML approach along with the data, and explainability considers the model, the data, and also human involvement (Roscher et al., 2020).

> ***Transparency****: if the model by itself is understandable, the model is considered to be transparent (Barredo Arrieta et al., 2020). Some transparent models include tree models and linear regression. A counterfactual example of non-transparency is the "black box." Roscher et al. (2020) further divided transparency into model transparency, design transparency, and algorithmic transparency.*
>
> ***Interpretability:*** *The aim of interpretability is to explain or provide the meaning of some of the properties of the AI systems in a human-understandable way (Barredo Arrieta et al., 2020; Roscher et al., 2020). Transparency can assist with interpretability, but the interpretability itself cannot achieve explainability if context information is not provided.*
>
> ***Explainability: explainability*** *refers to the capacity of explanation–usually in the format of an interface between humans and a decision-maker—to serve as an accurate representation of the decision-making process while remaining comprehensible to humans (Barredo Arrieta et al., 2020). In this case, explainability rarely can be achieved through pure algorithms (Roscher et al., 2020). A collection of interpretations is not sufficient either without contextual information derived from domain knowledge and task objectives (Roscher et al., 2020).*

In this dissertation, I use explainability as an overarching term to describe the interface and models that aim to rationalize the AI systems' decision-making processes.

2.3.2 Explainable Models

The advancement of ML techniques has increased model performance and accuracy. However, a tension between ML performance and model explainability has been found: methods with the highest performance (e.g., deep learning) are the least explainable ones (Gunning & Aha, 2019). Thus, the initial attempts for XAI are to explain complex models on how they reach certain decisions or outcomes and how the input variables lead to the outcomes. In other words, these existing XAI methods target data scientists or developers to help them understand the models and thus, improve model efficiency and ensure model accuracy (Barredo Arrieta et al., 2020). There are different types of XAI methods for post-hoc explainability when the complex models themselves are not transparent. These methods can be divided into local explanation and global explanation. Local explanations provide interpretability towards an individual or a small part of predictions, while global explanations present aggregated and ranked contributions of input variables for the entire model prediction. For each explanation type, there are model-agnostic and model-specific techniques. Model-agnostic methods do not require information related to the model internals and can be applied to any model (Barredo Arrieta et al., 2020; Holzinger et al., 2022). On the contrary, model-specific techniques can be applied to a single or a group of ML techniques that share similar characteristics (e.g., models that do not have layered structures) (Barredo Arrieta et al., 2020).

Another set of target users of XAI methods is the domain experts or the users of the model, who seek explanations to help them trust the model and gain scientific knowledge from the model (Barredo Arrieta et al., 2020). For instance, medical doctors using AI-powered systems that analyze the images are interested in knowing the system's accuracy and understanding how the system reaches a certain diagnosis. Situated in the AV context, the drivers are interested in knowing why the AV system acts in a certain way (e.g., slow down at an intersection due to pedestrian crossing) or why it proposes a specific maneuver. However, not many XAI methods are available for end users. The existing XAI methods for end users have not been fully explored and are not sufficient to develop their own taxonomies. The

present dissertation research aims to contribute to the exploration of XAI methods for end users in the driving context.

2.3.3 XAI from a Social Science Perspective

Although AI is mainly situated in the field of ML, the activity of explaining is social in nature. Considering that humans often treat computers as social actors (Nass et al., 1994) and endow human-like intelligent agents with human capabilities (Waytz et al., 2010, 2014), designing explanations provided by AI systems can refer to how humans explain their decisions and behaviors to each other. Miller (2019) reviewed philosophical, cognitive, and social foundations of explanation and highlighted four major findings that can enlighten the development of XAI. The following contents elaborate on each finding.

Finding 1: Explanations are contrastive. Josephson and Josephson (1996) defined explanation as "an assignment of causal responsibility." Regarding causality, the research found that people typically do not "explain the causes for an event per se, but explain the cause of an event relative to some other event that did not occur" (Miller, 2019). In other words, people explain in the form of "Why P happened rather than Q?" – this is called contrastive explanation.

Finding 2: Explanations are selected in a biased manner. People rarely expect an explanation that covers all aspects of the causes of events. Instead, they select one or two causes, a selection process that is largely influenced by cognitive biases. Confirmation bias is a good example of how people select evidence to support their beliefs. In the automated vehicle setting, the selection of explanations is critical to match drivers' information gaps and needs.

Finding 3: Probabilities don't matter and are not effective. Although likelihood and probabilities are important in hard science and making conclusions, they are not of great importance when it comes to explanations. Previous research found that numbers such as confidence level were less meaningful when explaining the automation system's decision-making confidence, and it did not provide practical suggestions on how users could handle the situation either (Peintner et al., 2022).

Finding 4: Explanations are social. Explanation in the social context is about transferring knowledge from explainer to explainee. Explanations do not need to be in the format of natural language, but they have to take social interactions into consideration, which means that the knowledge transfer is in the format of conversation and interaction and should take both the explainer's and the explainee's beliefs into account.

These findings, although abstract, have insights into how to design proper explanations for social goods and guarantee the effectiveness of explanations under the context of social interaction.

2.3.4 Metrics for Explainable AI

Up until this section, Section 2.3 has introduced several XAI models and guidelines. How to efficiently and properly measure the XAI system and whether it is suitable for the user tasks is also critical for the development and advancement of the field of XAI systems. Hoffman et al. (2018) proposed several metrics and corresponding measurement instruments to evaluate XAI, which has been officially published recently in Hoffman et al. (2023). Hoffman et al. (2023) presented the conceptual model of explaining situated in the XAI context, which pointed out the six key factors to be evaluated on an XAI system: **explanation goodness**, **explanation satisfaction**, **user mental model evaluation**, **curiosity**, **user trust**, and **human-XAI performance**. It is noted that although curiosity is not present in the Figure, it is an important construct mentioned in the metrics.

Explanation goodness and **satisfaction** basically refer to a similar construct—the quality of an explanation—but from a different perspective. Explanation goodness evaluates an explanation from the system designer's perspective, while explanation satisfaction evaluates an explanation from a user's perspective (Hoffman et al., 2023). In essence, users seek certain explanations on top of what they have already known to fulfill their goals. Hoffman et al. (2023) developed an Explanation Goodness checklist to help researchers and developers assess the *priori* goodness of an XAI system. On the contrary, the explanation satisfaction is a contextualized *posteriori* evaluation of the XAI system from the user, defined as the "degree to which users feel they sufficiently understand the AI system, or the process being

explained to them" (Hoffman et al., 2023). Key factors included in the Explanation Satisfaction scale are understandability, feeling of satisfaction, sufficiency of detail, completeness, usefulness, accuracy, and trustworthiness (Hoffman et al., 2023; Muir & Moray, 1996). For the explainable HMIs situated in AVs, the Explanation Goodness checklist can be used to assess the interface at the early cycle of the design activity, and the Explanation Satisfaction scale can be distributed to participants in the testing procedure. However, it is critical that driving, compared to other AI-assisted tasks, can be time-sensitive and safety-critical. Thus, a specific goal-oriented task analysis (GOAT) can be helpful in developing explainable interfaces and revising the checklist and scale as needed.

Users can still be satisfied with a system, while their mental model—understanding of the AI system—can be flawed. Thus, **measuring users' mental models** periodically while they are constantly interacting with the XAI system is critical. Methods that can elicit mental models include think-aloud tasks, structured interviews, and diagramming tasks. After collecting the user's mental model and giving an explanation, the mental model can be evaluated by proposition analysis for its correctness, comprehensiveness, coherence, and usefulness (Hoffman et al., 2023). In empirical studies researching drivers' understanding of driving automation systems, users' mental models are typically assessed at multiple time points (e.g., post-training, post-drive) (Noble, 2020).

Curiosity is an important factor in XAI because users seek explanations out of curiosity, and explanations can sometimes suppress curiosity and reinforce inaccurate mental models (Hoffman et al., 2023). When users realize that there is a gap between their current status and knowledge pool, curiosity may or may not come in to seek explanations, depending on user personalities. However, should users seek explanations, improperly framed explanations can also easily overwhelm or confuse users, which suppresses their curiosity for further explanation inquiries. There are different dimensions and theories about curiosity and how to measure it. Based on the knowledge gap curiosity, Hoffman et al. (2023) developed a curiosity checklist to measure curiosity situated in the XAI context. The explanations in AVs are often one-way, meaning explanations are provided by the system, and user inquiries are not taken (Phillips et al., 2021),

which limits user autonomy and information seeking. Thus, it is an urgent to explore two-way explanations that allow user interaction during the explanation process. Curiosity will be more easily and feasibly measured in two-way explainable systems.

User trust is a constantly discussed and explored construct in human-AI interaction. Fostering appropriate trust can prevent system misuse and disuse (Parasuraman & Riley, 1997). There are multiple definitions of trust. In the present dissertation work, I adopt the definition from Lee and See (2004) and define trust as "the attitude that an agent will help achieve an individual's goals in a situation characterized by uncertainty and vulnerability." The users voluntarily put themselves in a vulnerable situation by initiating the usage of AI systems, whose capabilities and limitations are often uncertain to users. Explanation plays a vital role in reducing uncertainty and facilitating appropriate trust. Overtrusting the AI systems can increase vulnerability as the users fade out the tasks and hand control and decision-making over to the automation system. Trust is also a dynamically changing construct as users interact with the system repeatedly. Thus, actively monitoring changes in trust as the explanations come in can help understand the explanation quality as well. Driving an AV on the road undoubtedly puts users in a more vulnerable situation with more uncertainties. Failures in AV operation can have more severe consequences when other dynamic objects are present on the road. The unintended consequence of overtrust is more critical compared to distrust. But eventually, with the popularization of autonomous driving, mistrust can incur greater issues in terms of technology acceptance and usage. Quantifying trust is the first step before calibrating trust. There have been different scales developed to measure trust. Hoffman et al. (2023) deconstructed the existing trust scales and reconstructed a scale that is mostly suitable for the XAI context.

Task Performance. Finally, the ultimate goal of the user seeking explanations is to achieve a certain task. Thus, measuring task performance should not be overlooked when evaluating an XAI system. The metrics for task performance are typically tailored based on the task. For AVs equipped with Level 1-3 automation systems, drivers are still required to perform certain driving tasks. Driving performance,

especially the reaction time to safety-critical events when prompted by the XAI system, is valuable in indicating the quality of human-automation interaction as a consequence of well or poorly designed XAI.

2.4 In-Vehicle Explainable Systems

Although no formal taxonomies or guidelines exist for XAI methods for end users, car manufacturers and research scientists have incorporated explanations into AV systems to facilitate human-automation interaction. The following sections introduce some in-vehicle explainable systems that are commercially available and then present explainable system prototypes that have been tested in the research setting.

2.4.1 Benchmarking analysis with AVs on the market

To the date of this dissertation, the vehicles available on the market are only equipped with driving automation systems that qualify as Level 1 or 2. Those active driving assistant (ADA) systems refer to the simultaneous engagement of both adaptive cruise control (ACC) and lane centering assistant (LCA). Some vehicles are also equipped with driver monitoring systems that supervise driver states and intervene if necessary. The Consumer Reports® (CR) released an investigation on 12 ADA systems from different automakers regarding their *capabilities and performance, keeping driver engaged, ease of use, clear when safe to use* (system status), and (reaction to) *unresponsive drivers* (Monticello, 2023). The *capabilities and performance* dimension refers to the automation system performance, which largely relies on the technology fineness. The *ease of use* is an overarching evaluation of whether the system has "simple controls, clear displays, and good feedback regarding the system's status" (Monticello, 2023). The *clear when safe to use* and *unresponsive drivers* fall into the functionality categories of indicating AV functional status and monitoring and intervening driver status mentioned in Section 2.1, respectively.

Based on the same report, Ford's BlueCruise outperformed the other ADA systems and received the highest score in *keeping driver engaged* (9/10) and *clear when safe to use* (9/10), followed by Chevrolet/GMC/Cadillac's Super Cruise feature that scored 8/10 in both *keeping driver engaged* and *clear when safe to* use, while the remaining 10 systems only received a maximum of 4/10 rating on the *clear when safe to use* (Monticello, 2023).

The findings from this report indicate the need for improving the ADA system's explainability, specifically in indicating the current AV status, considering all ADA systems can only be operated under certain circumstances with current technologies. However, it has to be admitted that such feedback may still rely on certain hardware requirements to be able to inform drivers on whether they should or should not activate the feature. For example, BlueCruise and Super Cruise both adopt GPS-based geo-fencing to ensure that accurate information regarding the system's operational domain can be communicated. Even though the vehicle was not able to incorporate geo-fencing, speeding, number of lanes, weather conditions can still be used to constrain drivers in situations where the ADA systems are not safe to use.

2.4.2 Prototypes evaluated in the research setting

Research scientists have also made attempts to understand what explanations drivers need in AVs. Wiegand et al. (2020) conducted a thematic analysis of real-world experiences and identified 17 unexpected vehicle behaviors. Among these 17 behaviors, 9 are situations of vehicles stopping abruptly without a good reason, 6 are about inappropriate speed choice or distance maintained, and the remaining 2 are about unclear interaction with other road users and unnecessary lane change. All these situations are directly related to driving conditions, which is also the main focus of empirical studies on developing and evaluating explainable interfaces. Shen et al. (2022) conducted an online survey to understand the explanation necessities given a driving scenario. Their findings indicate that near-crash, lane-changing, and slowing down are three primary scenarios with high explanation necessity, which aligns with the findings from Wiegand et al. (2020).

Koo et al. (2015) first proposed two different components of explanations: the "how" message and the "why" message. The "how" message explains how the AV is acting, which is later termed "what" information. **To be consistent, I use "what" information to refer to explanations regarding the AV's current or future action.** The "why" message provides situational information underlying the AV action.

To understand the effectiveness of explanations in SAE Levels 2-5 AVs, Q. Zhang et al. (2021) conducted a review of the explanation contents and timing and their effectiveness on driver outcomes, specifically on driver trust. The findings indicate that "why"-only explanations consistently promote user experience, while "what"-only explanations deteriorate driver performance. The combination of these two pieces of information showed mixed results in terms of emotional outcomes and driving performance. In terms of explanation timing, explanations before AV action resulted in positive emotional outcomes and can reduce driver anxiety and workload, while explanations after AV action did not provide benefits in human-automation interaction in terms of user trust and preference. While results from Q. Zhang et al. (2021) provide an overview picture of the explanation contents and timing, the explanation format and effectiveness can differ depending on the communication modalities. The following paragraphs present some findings on explanation modalities, specifically visual and auditory displays.

2.4.2.1 Explainable Visual Displays.

In addition to presenting information through traditional instrument panels, research studies have explored other visual display options and their effectiveness in delivering explanations. For example, **augmented reality (AR)** and **ambient light strips** are two dominant emerging displays evaluated in the existing research under different traffic scenarios. Colley et al. (2021) proposed four types of visual display concepts that communicated vehicle intention under unexpected pedestrian crossing cases: heads-up-display (HUD) with text messages, AR with text messages, light strip, and highlighted prioritized traffic elements (e.g., pedestrian, other vehicles) with text messages. The concepts consisting of text messages all included both "what" and "why" information. Results indicated that these visual display concepts increased the perceived safety, trust, and usability of the AV, but their effectiveness was not significantly different. Similarly, to address the anxiety of drivers when the SAE Level 5 AV approached crowded intersections, Colley et al. (2022) evaluated nine visual feedback types dependent on the AR windshield display (WSD) and found that user satisfaction was in general low, but feedback that blocked out the tension of the intersections was preferred. Research has explored the benefits of using ambient

light displays in SAE Level 5 AVs. Compared to the lights that only indicated the potential conflicts in the traveling route, the lights with additional vehicle trajectory information were claimed to have a higher value for users (Löcken et al., 2020). Recent work-in-progress research also adopted the "what"/"why" framework in an AR WSD for highly automated vehicles (Manger et al., 2023) and found that the "what" and "why" information combined led to the highest situational trust score and information processing awareness score, compared to no information and "what"-only information conditions.

Research also compared the effectiveness of WSDs and ambient lights when drivers were engaged in non-driving-related tasks (NDRTs). In terms of indicating vehicle action (e.g., braking, accelerating), ambient lights were perceived to have a better user experience and received higher preference compared to WSDs, although these two types of visual displays both facilitated trust and did not interfere with NDRTs (Dandekar et al., 2022). Research also explored HMIs supporting cooperative driving to assist drivers in making decisions under critical circumstances. Peintner et al. (2022) explored an HMI concept using HUD to provide drivers with recommended actions and the confidence level of the automation system for a pedestrian crossing scenario. This concept provided drivers with decision-making autonomy and supporting evidence, which is advocated for the evaluative AI that aligns with the cognitive decision-making process (Miller, 2023). However, confidence level did not add benefits to HMIs in assisting drivers but might introduce confusion as indicated by delayed decision time (Peintner et al., 2022). This finding aligns with the social perspectives of XAI that the numbers are not effective in explanations.

2.4.2.2 Explainable Auditory Displays.

Auditory displays are also extensively researched to provide explanations, considering their higher information capacity and lower interference with driving tasks. In addition to the "what"/"why" framework (Koo et al., 2015), the situation awareness (SA)-based framework was also developed to deliver speech messages (Avetisyan et al., 2022).

The original research attempt investigating the "what"/"why" framework found that, in a semi-autonomous driving condition, "what" information resulted in the worst driving performance in terms of road edge excursion, while the "why" information was preferred and yielded better driving performance (Koo et al., 2015). The combination of "what" and "why" information resulted in the safest driving behavior but also introduced negative emotional outcomes. Since then, it seems that providing explanations with a combination of "what" and "why" messages have been widely adopted in research. Combining auditory cues with a simulated visual display that mimicked the current HMI solution for adaptive cruise control with steering assistance, Yannick et al. (2017) compared the earcon warning with and without additional speech explanation ("what" + "why"). The auditory cue with both earcon warning and speech explanation received higher ratings on trust and usability. Du et al. (2019) also explored the effectiveness of the combination of messages in a simulated SAE Level 4 AV, but they mainly focused on the explanation timing (i.e., before and after AV action) and the degree of autonomy (i.e., whether drivers had the autonomy to make decisions on AV actions). Results indicated that explaining before the AV action yielded the highest trust ratings; drivers ranked explanations that were given before taking action and explanations that requested driver permission as higher in terms of driver trust (Du et al., 2019).

The situation awareness (SA) level-based explanation framework considers three levels of information processing established by Endsley (1995): Level 1 – Perception, Level 2 – Comprehension, and Level 3 – Projection. Based on this model, Avetisyan et al. (2022) constructed three levels of explanations with increasing information: Level 1 (SA L1 condition), Levels 1 + 2 (SA L2 condition), and Levels 1 + 2 + 3 (SA L3 condition). This explanation framework was examined on its impact on driver perception in highly automated vehicles through an online study. While SA L2 explanations received the highest situational trust, they also yielded a higher cognitive load. The explanation satisfaction also depends on the modality: the combination of visual and auditory displays was preferred for SA L3 explanations, while the visual-only display was preferred for SA L1 and SA L2 explanations. Similarly, Zang (2023) developed an explanation system based on the situation awareness-based agent transparency (SAT) model

(Chen et al., 2017) and evaluated L1, L2, and L3 explanations in a driving simulator study under both low and high system reliability conditions. The author found interaction effects between explanation types and system reliability that with L2 explanations, drivers perceived higher cognitive trust and lower workload when the system reliability was high compared to when it was low (Zang, 2023)

2.4.2.3 Explanation Effectiveness Compared across Communication Modalities.

To compare explanations using different unimodal feedback types, Shneider et al. (2021) evaluated shuttle riders' experience and preference toward five types of displays: light strip, earcon, visualization (similar to those instrument displays), text, and vibration. Light strips and visualization yielded a better user experience compared to no feedback condition (Schneider et al., 2021).

Based on the existing evidence, emerging visual displays have become the trending solutions for highly automated vehicles (SAE Levels 4 and 5), while auditory displays or tactile displays are more considered in conditionally automated vehicles. However, although researchers have explored different design alternatives, their focus remains on the general user experience. The explanation necessity and the underlying structure for explanations are not well considered, which leads to overall low user satisfaction found in Colley et al. (2022) and decreased system acceptance found in Graefe et al. (2022).

Although it seems that multiple attempts have been made to understand the effectiveness of different types of explanations, the fundamental question about what constitutes a "good" explanation is not well answered from the human understanding aspect.

2.5 Present Dissertation

The primary goal of this dissertation is to identify explanation needs in SAE Level 3 AVs and propose explainable in-vehicle multimodal displays to fulfill the explanation needs. In particular, this dissertation seeks to achieve the following objectives through a series of three sets of studies:

Chapter 3 Examining In-vehicle Intelligent Agents in Driver-Automation Interaction

- Explore the design characteristics for in-vehicle intelligent agents (IVIAs).
- Explore driver perception and performance under the influence of different IVIA designs in SAE Level 3 AVs

Chapter 4 Requirements Gathering: Explainable Information Systems in Automated Vehicles

- Gather user requirements in terms of their needs for explanation in AVs.
- Outline explainable interface design directions by consulting domain experts and usability engineers.

Chapter 5 Study 5: Designing and Evaluating Explainable In-vehicle Intelligent Agents

- Propose explainable interface alternatives based on the findings in Chapters 3 and 4 and existing theories.
- Examine and compare the effectiveness of the proposed interfaces through driving simulator studies.

CHAPTER 3 EXAMINING IN-VEHICLE INTELLIGENT AGENTS IN DRIVER-AUTOMATION INTERACTION

In-vehicle intelligent agents (IVIAs) can be a promising type of artificial intelligence (AI) system that provides explanations for the reasoning process and corresponding actions of automated vehicles (AVs). IVIAs are expected to proactively engage in driving tasks on a wider spectrum to secure road safety, supporting drivers in both driving-related and non-driving-related activities (Wang, Hock, et al., 2021).

IVIAs can provide vehicle status- and road condition-related information. This information is critical to constructing explainable automation systems to help drivers establish appropriate mental models, calibrate their trust towards automated vehicles (AVs), and facilitate technology acceptance. In conditionally automated vehicles, drivers are required to take over the control from the automation systems and intervene in critical events when AVs reach their operational limits. This automation-to-driver authority transition process is called a takeover process (McDonald et al., 2019). The smoothness and safety of the takeover process are primarily determined by the timeliness of the takeover requests (TORs) and the effectiveness of the driver intervention (McDonald et al., 2019). IVIAs in such a setting are expected to support driving-related tasks as one of their functional objectives to release drivers' burden (S. C. Lee & Jeon, 2022). If designed properly, IVIAs can play active roles in providing timely TORs and assisting drivers in negotiating safety-critical events.

Existing research has investigated the design components of IVIAs that contribute to the timeliness of TORs. Simply adding IVIAs increased the likelihood of drivers making timely reactions (Mahajan et al., 2021a). The time-to-collision at the time of the TOR (i.e., lead time), also significantly impacts the driver reaction time as well as the lateral and longitudinal post-takeover driver intervention (McDonald et al., 2019). Generally, longer lead times are associated with longer takeover times (McDonald et al., 2019). In addition to the temporal variable, varying signal words, the tone, or the loudness of speech can convey

different levels of perceived urgency, which further influences takeover reaction times (Politis et al., 2015; Roche & Brandenburg, 2020; Wong et al., 2019).

IVIAs can also assist drivers in strategizing maneuvers and controls because they are capable of carrying versatile information. Effective driver controls are guarded by appropriate situation awareness (SA) in drivers, which is largely compromised in conditional AVs. Drivers in conditional AVs—as passive monitors—have declined alertness caused by task disengagement, low workload, or passive fatigue (Mahajan et al., 2021a; Vogelpohl et al., 2019). Vehicle status- and road condition-related information provided by IVIAs can keep drivers in the loop, improve their SA, and prompt future actions (Hester et al., 2017; Mahajan et al., 2021a, 2021b; Nees et al., 2016). For instance, the *"how*" message announces the vehicle's current action, and the "*why*" message explains the reason for vehicle decisions (Koo et al., 2015), while the "*what will*" message provides further recommendations in reaction to the scenario (Du et al., 2021). The "*what will*" message was perceived as more useful and easier to use compared to others in conditional AVs (Du et al., 2021). These semantic contents are important to improve drivers' SA and are critical to communicating automation capabilities and limitations, helping calibrate trust and avoid takeover failure due to system misuse (J. D. Lee & See, 2004; Parasuraman & Riley, 1997). Social attributes (e.g., speech style, embodiment) of IVIAs are also beneficial in trust calibration in conditional AVs, which can further impact the effectiveness of driver intervention.

Exactly due to their versatility, IVIAs should be carefully designed. The first part of this dissertation investigates the effectiveness of using IVIAs with different design features in assisting drivers with driving tasks. This Chapter is structured in the following. Section 3.1 presents a participatory design workshop that collected insights from experts and practitioners in the field of AVs to direct the design guidelines of IVIAs. Section 3.2 presents a driving simulator study that examined the influence of IVIAs with different embodiment conditions and speech styles in drivers' perception and takeover performance in the conditionally automated driving condition – one of three use cases discussed in Study 1.

3.1 Study 1: Workshops on Characterizing and Designing IVIAs

As mentioned above, previous research has provided valuable insights into the features of IVIAs and the resulting user perception. However, characteristics that uniquely scope IVIAs and distinguish them from at-home voice assistants remain untouched. To further characterize and optimize IAs situated in the driving contexts, I, as the main organizer, hosted two workshops on characterizing and designing in-vehicle intelligent agents at the 13th and 14th International ACM Conference on Automotive User Interfaces (AutoUI 2021, AutoUI 2022) (Wang, Hock, et al., 2021; Wang, Park, et al., 2022).

The two-day workshop held at AutoUI 2021 aimed to gather the opinions of the experts and practitioners in the field of automated vehicles to scope the definition of agents, and then characterize IVIAs in terms of their specific functions and design features that differentiate them from at-home agents. An IVIA is comparable to Jarvis in the Marvel Universe, who is able to handle a various range of tasks even without user commands, including safety alerts and route planning. On the other hand, an at-home agent is comparable to Genie in Disney's animated feature film, Aladdin, who is capable of almost everything per request. In other words, IVAs are proactively engaged in driving tasks for monitoring and notifying road safety conditions, while at-home agents only passively respond when users ask for information or make a command. In addition to distinguishing IVIAs from at-home agents, the workshop also collected a cluster of functions carried by IVIAs under the three representative levels of automation conditions: Level 0 with no driving automation, Level 3 with conditional driving automation, and Level 5 with full driving automation, considering the dynamic function allocation between an agent and a driver as the level of automation increases. In total, 30 participants from eight countries occupied in both academia and industry attended the first workshop.

The half-day workshop held at AutoUI 2022, as a continuous discussion regarding the IVIA design considerations, aimed to provide hands-on experiences for the attendees to design their own IVIA. The main objectives of this workshop are to (1) integrate a list of design variables and characteristics of IVIAs, (2) to develop IVIA prototypes especially considering factors of appearances and voices, and (3)

to investigate user preferences towards IVIA prototypes under different user scenarios. In total, 10 participants from four countries occupied in both academia and industry attended the second workshop.

3.1.1 Methods and Activities

As participatory design methods, these two workshops adopted expert focus groups and design activities. Participatory design allows users and stakeholders to actively engage in the design process, which ensures that designers are addressing user problems and grounding design decisions from target stakeholders (Sharp et al., 2019). While it was challenging to involve the direct end-users in the design process for future technologies that do not exist yet, the experts and practitioners in the field of automated vehicles (AVs) could serve as subject matter experts. Thus, their opinions were valuable in addressing the challenges associated with designing IVIAs situated in future AVs.

3.1.1.1 Focus Group

The Focus Group method was only used in the first workshop at AutoUI 2021. On the first day of the workshop, the attendees formed five groups and engaged in focus group discussions. The groups exchanged their ideas on four topic questions in sequence.

1 *What makes an agent an agent?* The discussion on the first question aimed to establish a definition of agents. The factors that define an agent are referred to as agent **attributes** in the following sections.

2 *What are the special characteristics of IVIAs?* After defining agent attributes, the section describes both the unique functions and design features of the agents that differentiate them as at-home agents or IVIAs. **Functions** refer to the purpose of the agent, in other words, what the agent is used for. For instance, functions of IVIAs include but are not limited to monitoring the traffic conditions and informing drivers properly. **Design features** specify both the visual and auditory specifications of agents, such as agent appearance, agent voice, and agent facial expressions (if any). In addition to these objective features, perceived features such as agent attitudes and user trust are discussed.

Functions and design features together characterize the agent into different categories: at-home agents or IVIAs.

3 *How do drivers and passengers perceive or distinguish these two categories of agents?* With specified characteristics for IVIAs, the discussion further extended to how users distinguish IVIAs from at-home agents depending on variations in functions and design features.

4 *What are users' preferences toward the form of agents?* The last part discusses whether users prefer a one-fits-all agent that serves users both at home and in the vehicle or multiple agents specialized in different situations.

3.1.1.2 Design Activity

Both workshops involved design activities that allowed attendees to freely design their own IVIAs with given constraints and instructions.

For AutoUI 2021, on the second day of the workshop, returning attendees formed three groups to brainstorm characteristics for an agent to serve in each of the three levels of automation conditions (Level 0, Level 3, and Level 5). The group first determined agent functions and then considered design features.

For AutoUI 2022, after briefly reporting the results from the workshop at AutoUI 2021, attendees were divided into two groups to discuss a desired use case scenario and then brainstorm a dialogue between the driver (or the rider) and the agent.

3.1.2 Results: Focus Group Discussion

Affinity diagrams, the simplest way to reveal common themes across all idea contributors (Holtzblatt & Beyer, 2017c), were used to organize attendees' discussion points brought out in the first-day topic discussion session.

3.1.2.1 Defining Agents

Both agent attributes and characteristics were discussed to define agents (Figure 3.1-1). User-centered action and agent autonomy are two key attributes that qualify an interactive entity as an agent. An agent performing user-centered action will always attend to the user and pay attention to the user's status. Even though the agent cannot provide an instant response to a request, feedback on the system's operating status should be provided. In addition to always attending to users, an agent can understand user needs and requirements and can consider user preferences when analyzing and responding to user requests. In fact, user-centered action is an attribute shared by all human-machine interfaces. In terms of autonomy, an agent should retain a certain level, if not all, of control over the vehicle automation system in terms of decision-making to prioritize tasks based on its objectives.

In addition to the attributes, agents' functions and design features are two critical characteristics that identify an agent. The overarching function of an agent is to assist users to perform their tasks, either driving-related tasks or non-driving related tasks (NDRTs). The design features can be further categorized into objective features and perceived features. The objective features typically refer to the properties of an agent, such as the speech style, voice gender, speaking speed, etc. The perceived features refer to the subjective evaluations of agents that are perceived from humans, such as trust, competence, warmth, etc. These perceptions, affected by the agent outputs defined by their functions and objective features, are important design considerations to assess agent quality. A well-designed agent can present information unknown or unclear to users, such as system limitations or external information hidden from the user, improving system transparency and fostering user trust (Chen et al., 2017; Wright et al., 2017).

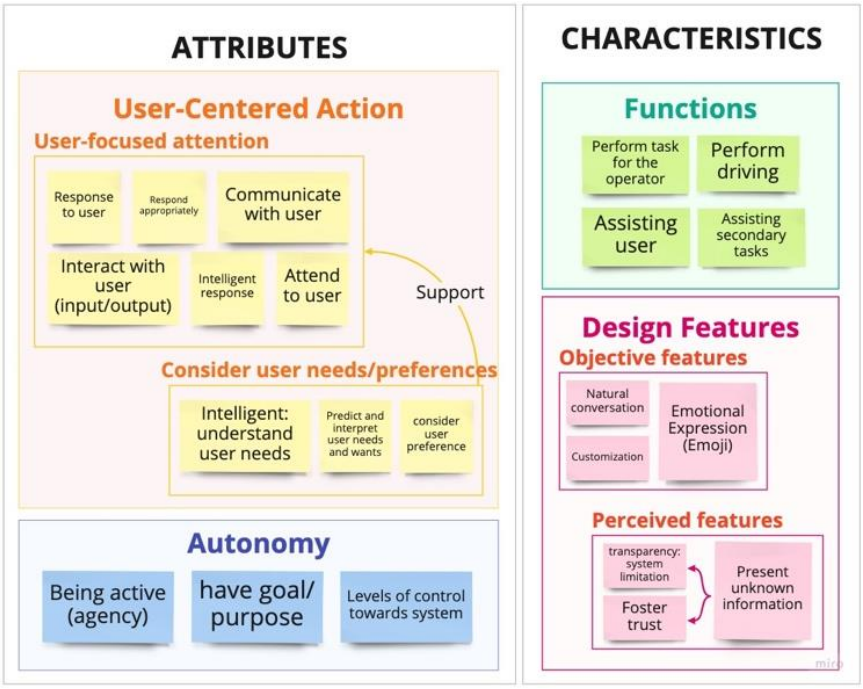

Figure 3.1-1 Affinity diagram on defining agents.

3.1.2.2 Characteristics of IVIAs

As identified in the previous section, functions and design features are able to characterize agents (Figure 3.1-2). In the driving context, IVIAs can assist users with their primary tasks, such as driving, or support them with their NDRTs (e.g., interacting with the infotainment system) to ensure the primary task performance. In addition to functions directly related to driving support, IVIAs also have the potential to monitor and intervene in driver states such as their emotions (Jeon, 2015; Jeon et al., 2014; Jeon & Walker, 2011), drowsiness (Ghizlene et al., 2019), and boredom (Samrose et al., 2020) that might influence driving performance.

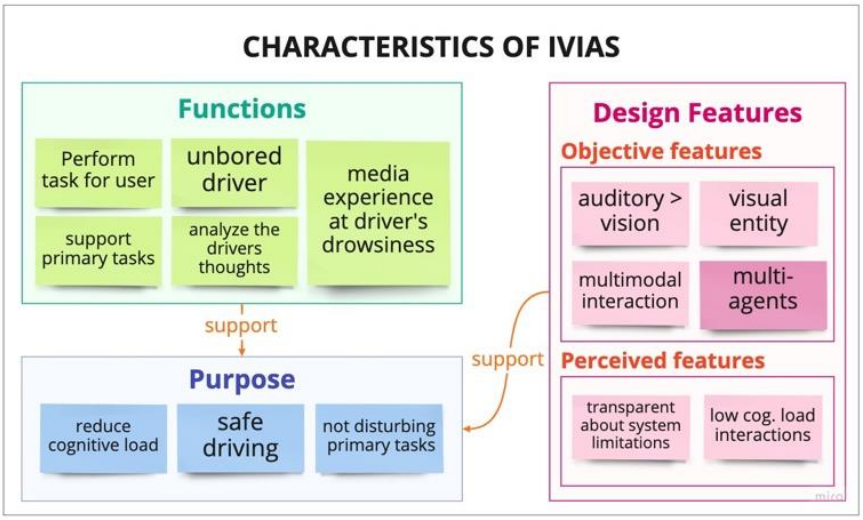

Figure 3.1-2 Affinity diagram on characteristics of IVIAs

Design features were also discussed heavily as special characteristics of IVIAs. Multimodal interaction was recommended to support driving as a visually demanding task, especially utilizing the benefits of the auditory modality to convey messages. The necessity of a visual entity for the presence of IVIAs was also discussed. It was argued that a visual presentation could be good for reference purposes, but it can also be distracting. Simplified but effective interaction is also advocated to further reduce the cognitive demand required for communicating with IVIAs. Other than features that support user tasks, features that facilitate user experience were also discussed as a potential for IVIAs but with less power of uniquely characterizing IVIAs. An example is the ability to customize IVIAs for multiple users. The system can take driver personalities into account to maximize user acceptance (M. Braun et al., 2019). Typically, the design of objective features can influence perceived features. For instance, a female voice was perceived with higher likability and competence in fully automated vehicles (Dong et al., 2020).

The group discussion also brought up a new research topic to be examined in the field of IVIAs. As IAs have become popular in our lives and integrated well into people's personal devices, multiple agents co-existing in the vehicle will be more common. The current knowledge available is not able to solve the problem concerning streamlining the interaction logic and resolving different agents' conflicting

responses. A concept paper has also pointed out this futuristic dilemma and embodied potentially challenging situations (S. C. Lee et al., 2021).

3.1.2.3 Genie vs. Jarvis

The discussion on perceiving and distinguishing at-home agents (e.g., Genie) and IVIAs (e.g., Jarvis) fell on functions and design features again (Figure 3.1-3). At-home agents and IVIAs perform different tasks in essence, with IVIAs more active in driving scenarios. Thus, users maintain corresponding expectations toward different agents. Agent voice is the most representative design feature for an agent and can differentiate one from another, aligning with the theory of computers as social actors (Nass et al., 1994).

The discussion came after distinguishing "Genie" and "Jarvis" was the issue associated with having multiple agents. Both information pieces input to and output from agents were discussed. Depending on the contexts, the agents can be initiated by voice commands at home with lower noise, button initiation in the vehicle where voice commands can be masked, or multiple-user scenarios can confuse the agent. Agent hierarchy is also situated in the context. The agent currently performing the task should be given the priority to interpret user needs and provide the appropriate response. If none of the agents is active when requested, an initiation action is required either through physical contact (e.g., button pressed) or a specific point of reference (e.g., "Hey, Siri"). Agent-to-agent communication was also brought up and discussed in the next section.

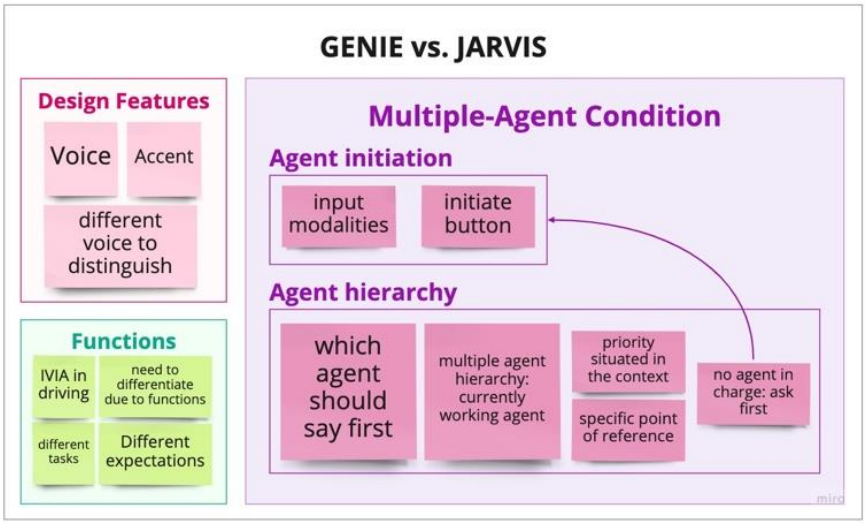

Figure 3.1-3 Affinity diagram on how to distinguish agents.

3.1.2.4 One Fits All vs. Specialized Multiples

Finally, the discussion on the preference towards the one-fits-all agent or multiple specialized agents concluded on the first day of the workshop. While it highly depends on the user context, use domain, and personal taste, most attendees favored specialized multiples and debated the advantages and disadvantages of both options (Figure 3.1-4).

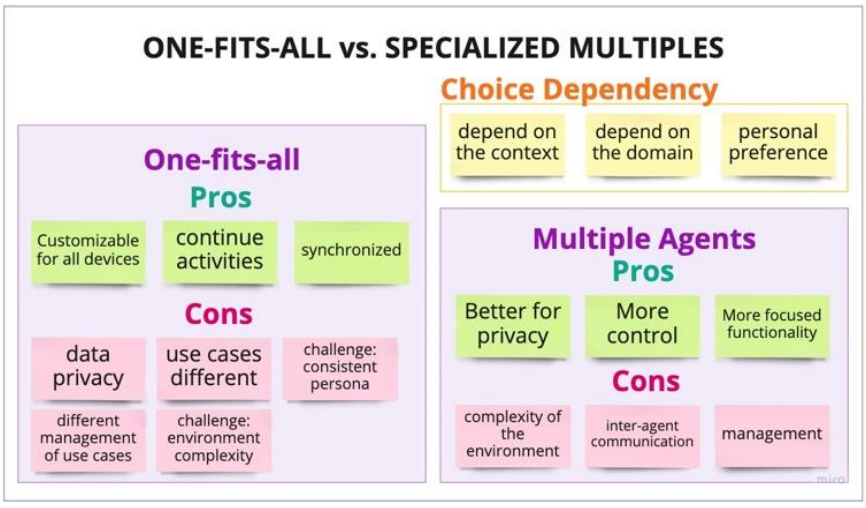

Figure 3.1-4 Affinity diagram on preferences towards agent type.

Although there are challenges associated with agent management and inter-agent communication, multiple specialized agents are more focused on delivering well-designed functions, tuning sophisticated

control, and providing better privacy. On the contrary, a one-fits-all design, thanks to the consistent persona established in a single system, allows synchronized information among different tasks and provides a streamlined experience in continuing the same activity across different user contexts. However, data privacy and agent complexity will become more of a concern. Further investigation into this trade-off is needed to optimize futuristic user experiences in the car.

3.1.3 Results: Design Considerations for IVIAs

Even though IVIAs are used in the same context (i.e., driving), they can play different roles in response to different automation levels. Thus, three key levels in the spectrum of automation levels (SAE International, 2021) were selected as representative use cases to fit IVIAs: Level 0 with no driving automation, Level 3 with conditional driving automation, and Level 5 with full driving automation. Three groups discussed the functions and design features of each IVIA fitting for each use case. Design features considered in the design activity section of the workshop included agent form factor (e.g., physical or virtual), agent appearance, speech characteristics, speech style, conversation methods, agent attitude, affective expression, body gesture, and other features proposed by attendees. The following sections summarize the design considerations for each use case, followed by general recommendations on agent design.

3.1.3.1 IVIAs in Level 0 Automation

IVIAs in Level 0 automation are designed to reduce drivers' cognitive demand in driving-related tasks and non-driving related tasks (NDRTs). Ideally, the agent is aware of the driver's state (e.g., tiredness, anger) and can adapt accordingly. Regarding design features, the voice-only form factor was advocated to situate in the vehicles with Level 0 automation to reduce visual distraction and avoid visual attention competition. Thus, the feedback on agent attention and operation (e.g., auditory cue) has to be properly designed to inform users about the system's status. Speech characteristics, speech style, and agent attitude can be customizable and are use-context dependent, while a one-fits-all default option should also be provided.

3.1.3.2 IVIAs in Level 3 Automation

In conditionally automated vehicles, as drivers are required to take over the control when the system requests, IVIAs should inform drivers about upcoming events and whether the system is able to handle the event. It is also important to provide explanations regarding current vehicle behaviors and their understanding of the current situations to establish trust and promote driver-agent interaction.

The design features fitted for IVIAs in Level 3 automation are slightly different from those designed for Level 0 automation. IVIAs in semi-automated driving can have a physical body seated on the dashboard. Other features are shared with agents in Level 0, especially the safety-driving-related design considerations such as driver state monitoring and intervening. The discussion on IVIAs in Level 3 automation also brought up an important point about keeping the agent consistent but also adaptive. Essentially, features that determine the vocal timbre—such as age and gender—should be constant once set up to ensure a consistent voice-identity perception and to distinguish IVIAs from other agents (Mathias & von Kriegstein, 2019), while the speech style and speech tone can vary depending on situations to convey additional information such as urgency (Lavan et al., 2019).

3.1.3.3 IVIAs in Level 5 Automation

IVIAs in fully autonomous driving are expected to be versatile to match the advanced technology applied to automation. IVIAs are expected to cover all aspects of the in-vehicle user experience. Thus, the design features are more user-oriented to provide personalized experiences.

Attendees also discussed another design feature that helps users perceive the agent as the same or refer to the agent properly in multiple-agent conditions. Besides agent voice, agent appearance can also identify and distinguish agents.

3.1.4 Results: User-Defined Multiple Agent Use Cases

To continue the topics from our first workshop, the attendees in our second workshop identified use cases that fall under the “One Fits All vs. Specialized Multiples” discussion theme. Two groups of attendees

brainstormed two use cases under different scenarios. One of the use cases was more generic about the IVIA design in highly automated vehicles, with a multiple-agent scenario. Another use case was specific under the riding-sharing context, where each passenger had their own agents.

3.1.4.1 Multiple Agent Use Case

One group of attendees advocated for two agents for driving tasks and NDRTs separately. In terms of the IVIA for driving tasks, this group advocates for the ambient light option without auditory display. A visual representative for the agent was not considered necessary but could cause potential distractions that compromised driving safety. In terms of NDRTs, the attendees advocated having customized avatars and disabling the ambient light when the passenger was resting in the vehicle instead of driving. The auditory display was, in general, not favored in this group, except for the extreme cases. Attendees in this group also had different tastes in customized avatars (Figure 3.1-5). Three attendees selected abstract representatives of the agent, and one of them used sound waves. Two attendees preferred figurative agents. One of them was the famous bear in Korea (Figure 3.1-5, top left), which indicated that cultural factors can also play a role in selecting customized avatars.

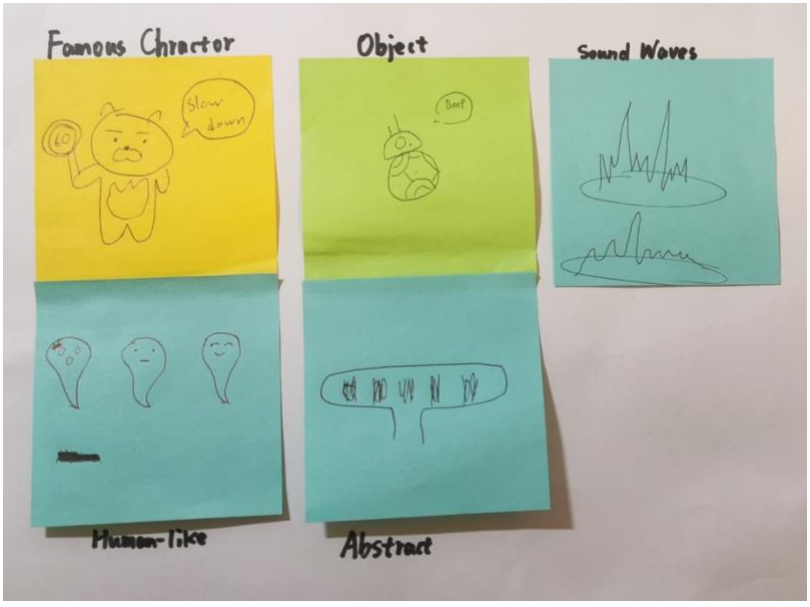

Figure 3.1-5 Customized avatars from Group 1 attendees.

3.1.4.2 Ride-sharing Use Case

Another group created a detailed dialogue that documented a ride-sharing use case where two passengers speaking different languages shared the same shuttle. In this context, each passenger had their own agent, but two agents could communicate with each other. The dialog happening throughout this ridesharing can be seen in Appendix A. It can be seen from the dialogue that multiple agents in the ridesharing scenarios are expected to be responsible for the following tasks: (1) taking user inputs for navigation or in-vehicle configuration if privacy is requested, (2) reminding users of their tasks, and (3) coordinating between passengers especially if multiple languages are spoken at the same time. In addition to these functions, agents in the ridesharing context are also expected to improve their sense of security (Schuß et al., 2022).

3.1.5 Discussion

Two consecutive workshops provided a comprehensive exploration of intelligent agents with the driving context. The discussion topics spanned from fundamental definitions of agents to specific design considerations for in-vehicle intelligent agents (IVIAs).

While user-centered action and autonomy represent crucial attributes for all types of intelligent agents, agents' specific functions and design features are largely dependent on the context in which they operate. Agents can be tasked with primary activities for users or assist with secondary tasks, depending on human-machine function allocation. Even within the same context, such as driving, an agent's functions can vary significantly based on the level of system automation. Therefore, it has been emphasized that clear communication of function allocation between humans and agents is essential in collaborative tasks (Lyons, 2013). Design features, like communication styles and capability of customization, are also context- and task-dependent. Both the functions and design features of agents contribute to serving the contextual purposes, and these two aspects are sometimes, but not always required to be aligned (Lyons, 2013; Sarigul et al., 2020). For instance, Hosseini et al. (2017) found that the people preferred a human-like robot with emotional expressions, indicating user preference for the matching between the designing feature (e.g., appearance) and the function (affective component).

Once agents were defined with their functions and design features, specific considerations within these two attributes were discussed. Through the discussion, proactivity emerged as a critical characteristic that sets IVIAs apart from at-home agents. Even when not directly queried by the driver, IVIAs proactively engage in tasks like monitoring vehicle states and evaluating road conditions to enhance driving safety. In contrast, state-of-the-art at-home agents primarily wait for user request passively. Subsequent discussions focused on well-designed features that could reduce drivers' cognitive demand, aligning with continuous research effort in IVIAs. As reviewed in Section 2.2.2, previous research has explored various design features to support driving tasks, NDRTs, or both, to reduce cognitive demands and improve driving performance. For instance, research has investigated different speech styles to improve takeover time using assertive style over the non-assertive style (Wong et al., 2019), reduce distractions when interacting with in-vehicle infotainment systems (Gaffar & Kouchak, 2018), or optimize the overall driving experience (Wang, Lee, et al., 2021).

While there is consensus on distinguishing IVIAs from at-home agents, a new debate has arisen about whether to have **a one-fits all agent** for all tasks or **multiple agents** dividing specific responsibilities, such as one for driving-related tasks and another for personal task management. Both advantages and disadvantages of each option have been discussed, and there is no consensus on the best choice. Instead, the decision should be based on the specific context, usage domain, or personal preference. This debate extended to the second workshop, where two user-generated use cases were proposed for a single user and a shared shuttle scenario, both constructed under highly automated driving conditions (Level 4 and above). These use cases emphasized the value of having multiple agents in terms of privacy and communication efficiency, potentially leading to an investigation of agent-agent interactions in addition to human-agent interactions. Existing evidence on multi-agent setups in autonomous driving suggests that having multiple agents responsible for different tasks can support trust calibration if designed properly (Premstaller et al., 2023). Within this discussion, one of the challenges is to distinguish among multiple agents. The discussion revealed that agent voice and appearance are features that identify or differentiate

agents. Especially for factors that contribute to the vocal timbre of agents, such as age and gender, they should remain constant once set up to ensure a consistent voice identity perception and distinguish IVIAs from other agents (Mathias & von Kriegstein, 2019). Other features such as speech style, tone, and pitch can vary based on the situation. Manipulating these objective features can also be helpful in inducing target perceived features such as perceived urgency (Lavan et al., 2019) under safety-critical situations.

Parasuraman et al. (2000) introduced the concept of different levels of human interaction with automation, which is valuable for establishing norms between drivers and driving automation systems. Findings from the design activity on the second day of the first workshop indicated that while IVIAs serving under varying levels of automaton serve different purposes and functions, design considerations remain similar, with minor adjustments to better align with those functions.

Across the two workshops, it is evident that while there are universal requirements for defining IVIAs, users tend to prefer IVIA customization, especially in highly automated vehicles. In contrast, having visual representations are less favored in vehicles with lower levels of automation, which aligned with my previous research findings under conditionally automated driving (Wang, Lee, et al., 2022).

Considering that the present dissertation only focuses on agents responsible for driving tasks, I will adopt a single agent set up and will not explore the multiple-agents option. Further research is needed to systematically explore the context and tasks required for the multiple-agents option.

The findings from the workshop discussions supports Lyons' theory on Human-Robot Interaction (Lyons, 2013), where "robot" in this context refers to any agents or systems that have some level of autonomy. Specifically, the workshop results revealed the importance of communicating agent purpose and their current tasks, which aligned with the intentional model and task model under robot-to-human transparency, respectively (Lyons, 2013). Then, the function allocation based on the level of automation and the agents' empathic features (e.g., emotion expression and driver emotion monitoring) matched with

the teamwork model and human state model under the robot-of-human transparency. The subsequent research activities will be guided by the findings from this workshop.

3.2 Study 2: Evaluating IVIAs in Conditionally Automated Vehicles

Study 2 aims to systematically evaluate the effects of speech style and embodiment—as two characteristics contributing to social attributes of IVIAs— and their interaction effect on driver perception and takeover performance in conditional AVs. Two representative speech styles—informative vs. conversational—were used to create divergent perceptions: the informative style sounds commanding due to its simplicity and directness, while the conversational one is more suggestive and can create a feeling of being cared for (J. D. Lee et al., 2017c). The influence of the absence or presence of a physical body (voice-only vs. robot) was examined to understand possible distraction introduced by embodied agents while also maximizing the anthropomorphism provided by a humanoid robot (Roesler et al., 2021; Złotowski et al., 2015). We adopted a within-subjects factorial design, attempting to detach the effects of two social attributes from each other and provide an unambiguous view of their influential mechanisms. Based on the existing evidence, we further hypothesize that:

H1: Drivers will prefer voice-only agents without visual distraction in conditional AVs over robot agents, while conversational speech style will also be preferred in this context to enhance driver experience and engagement.

H2: Drivers accompanied by conversational agents will have better performance, specifically:

H2.1: Drivers will have better SA.

H2.2: Drivers will demonstrate safer takeover performance.

The first study leads to the following unique contributions. First, shifting in driver experience, especially their preference towards IVIAs, when compared with IVIA design in full AVs, uncovers and supports the dynamic user needs and requirements as the levels of automation alter. Further, user perception of the driver-agent interaction provides insights on how to design agents to promote driver experiences in conditional AVs, laying the foundation for comparing IVIA designs and preferences across different levels of automation. Finally, findings on the influence of different types of IVIAs on the subsequent

takeover process can help with a balanced design between subjective preferences and unintended consequences.

3.2.1 Methods

3.2.1.1 Participants

Participant recruitment information was sent out after the study protocol was approved by the Virginia Tech's Institutional Review Board (IRB# 19-088). Twenty-four participants (7 females) aged between 19 and 33 years old (*Mean* = 23.12, *SD* = 4.49) with normal or corrected-to-normal vision participated in our study. All participants had valid driver's licenses and had an average driving experience of 5.93 years (*SD* = 3.37), with an average driving frequency of 4.67 days per week (*SD* = 2.22). Two participants had experience in partial AVs (i.e., Tesla full self-driving) before their participation.

3.2.1.2 Experimental apparatus and stimuli

We conducted a driving simulator study in a motion-based driving simulator (Nervtech™, Ljubljana, Slovenia), which consisted of three 48" displays that created a 120˚ horizontal field of view, an adjustable seat, a steering wheel, pedals for gas and brake, and surrounding sound equipment. Driving scenarios programmed in SCANeR studio were designed to simulate an SAE Level 3 Conditional Driving Automation (SAE International, 2021). We developed four driving scenarios, including straight and curved roads with traffic, traffic signals and signs, and other road users (e.g., pedestrians and other vehicles). The simulated ego vehicle had longitudinal and lateral control while navigating along a predefined route and handling limited road events such as stopping at a red light and crossing a controlled intersection. When the system reached its limitation (e.g., limited visibility due to weather, surprising event), a speech takeover request (TOR) along with a visual notification on the navigation panel would be issued to mandate the participant to take over control of the vehicle. The participant deactivated autonomous driving by either using a toggle attached to the steering wheel or pressing the brake. Upon exiting the takeover event zone, the system prompted the participant to reengage the automated mode (i.e., "Please reengage the auto-drive"). Each scenario consisted of four takeover events and four non-

takeover events, lasting approximately seven minutes. The route and order of events were different among the four scenarios to minimize learning effects.

Speech messages regarding road events and TORs used in this study were converted via the text-to-speech engine in Amazon Polly (name: Joanna, gender: female, nationality: USA). A humanoid robot, NAO (V6 standard edition, height: 22", width: 10.8"), was used under the embodied agent conditions. Figure 3.2-1 presents the experimental setup for the driving simulator and NAO. To capture participants' gaze fixation during the study, an eye-tracking device—Tobii Pro Glasses 2—with a sampling rate of 50 Hz was used.

Figure 3.2-1 Experimental setup with the NervTech driving simulator and Nao.

3.2.1.3 Experimental design

This study adopted a 2 (Speech style: informative vs. conversational) x 2 (Embodiment: voice-only vs. robot) within-subjects factorial design. Thus, four types of in-vehicle intelligent agents (IVIAs) were evaluated in this study: informative voice agent (IVA), informative robot agent (IRA), conversational voice agent (CVA), and conversational robot agent (CRA). Each participant was accompanied by all four agents in four different driving scenarios, respectively. The order of the agent conditions was counterbalanced across participants and with the matching scenarios. Thus, the same agent was not always used in the same scenario to avoid the scenario as a confounding variable.

The IVIAs issued TORs and provided information regarding road events (Table 3.2-1). All road events shared similar elements across four scenarios but were placed at different locations along the travel route to avoid learning effect. Thus, the difficulty of takeover events remained similar. While informative agents present information in a simple manner without additional information other than road events (e.g., "Exit ahead"), conversational agents communicate the message in a dialogue style (e.g., "We are entering a new road."). The TORs remained the same between informative and conversational styles to control the message length as a confounding variable that could impact information processing time and further influence the takeover reaction time under emergency situations. The length of takeover requests ranged from 2.41 to 3.00 seconds (*Mean* = 2.64, *SD* = 0.27) across four events. For other road events, the length of informative messages ranged from 0.86 to 1.31 seconds (*Mean* = 1.09, *SD* = 0.21), and the length of conversational messages was between 1.13 to 4.08 seconds (*Mean* = 1.97, *SD* = 1.39).

Table 3.2-1 Driving events and scripts in scenarios.

Event List	Informative Script List	Conversational Script List
Road construction	Take over immediately. Road construction ahead.	
Fog	Take over immediately. Fog ahead.	
Jaywalking	Take over immediately. Jaywalker ahead.	
Tunnel	Take over immediately. Tunnel ahead.	
Exit or enter a new road	Exit ahead.	We are entering a new road.
Waiting for a traffic signal	Red light ahead.	We are waiting for the signal to turn green.
Turning left/right	Turning left/right ahead.	We are turning left/right.
Two-way stop intersection	This is a two-way stop.	We've reached a two-way stop. We are waiting for other cars to go first.

The lead time was 4.5 seconds in this study. We selected a relatively limited time budget for the following reasons. First, our pilot study with a 7-second lead time for takeover events indicated a lower level of task difficulty, which led to performance degradation or passive fatigue due to boredom. Thus, we increased the task difficulty by limiting the time budget to keep participants' active engagement. Second, an empirical study showed a 4.5-second time budget led to a minimum crash rate and brake-to-maximum reaction time to speech warnings with a lead time shorter than 7 seconds (Y. Zhang et al., 2016).

3.2.1.4 Dependent measures and analysis

Both subjective and objective dependent measures were considered. Subjective measures included questionnaires to evaluate driver-agent interaction experience and driver preference. Objective measures included situation awareness, eye-tracking measures, and takeover performance. The following sections introduced the dependent measures separately and explained the analysis method afterward.

Subjective measures. Subjective measures included ratings from three questionnaires collecting driver perception on the accompanying agent: the modified Subjective Assessment of Speech System Interfaces (SASSI) (Hone & Graham, 2000) (the Habitability and Speed subscales were removed due to their irrelevance to our agent setting), and the Scale of Trust in Automated Systems (Jian et al., 2000). In addition, participants' preferences and reasons behind their first and least preferred agents were also asked. A two-way repeated-measures analysis of variance (ANOVA) was conducted to understand the influence of speech style and embodiment and their interaction effect on each factor of driver-agent interaction questionnaires. A Chi-square test for each preference rank was conducted to identify differences in preferred agents.

Situation awareness. Drivers' situation awareness (SA) was evaluated using the Situation Awareness Global Assessment Technique (SAGAT) (Endsley, 1988). To develop the SA queries, we conducted a Goal-Directed Task Analysis (GDTA) to identify the SA requirements (Endsley & Garland, 2000) needed for drivers under conditional automation to make decisions if a TOR was issued. With a list of SA requirements, six queries were constructed for each freeze point in the driving scenario, consisting of two queries for each level of SA: perception, comprehension, and projection (Endsley, 1995). Each query had four options with only one correct option. Participants' overall accuracy rate for each query was calculated. Queries with a lower than 25% accuracy rate (N = 2) of guess level were removed from further analysis. Then, the frequency of correctness (% correct) was calculated for each scenario. Because data from queries scored as correct or incorrect were binomial, the arcsine-root-square transformation was applied as a correction factor to allow the ANOVA tests (Cohen et al., 2013; Endsley, 2021).

Eye-tracking measures. The eye-tracking data were collected and stored in the eye-tracking device. We primarily focused on gaze fixation to identify any potential distraction due to introducing an embodied agent. Specifically, we calculated distraction fixation frequency and total distraction duration.

Gaze fixations on predefined areas of interest (AOIs) were identified in the Tobii Pro Lab software (v1.152) and classified into two primary categories: driving-related fixations and distraction fixations. Driving-related fixations included fixations to the road, other road users, road signs and signals, rear-view mirrors, and the instrument panel. Distraction fixations included fixations to NAO in robot agent conditions and personal devices. The distraction fixation frequency and total distraction duration were calculated using the formula below (Wang, Lee, et al., 2021; Zahabi & Kaber, 2018):

$$Distration\ Fixation\ Frequency = \frac{FixationCount_{distraction}}{FixationCount_{total}} \quad (1)$$

The total distraction duration was the sum of all distraction fixation duration in seconds. After the calculation, a two-way repeated-measures ANOVA was performed to identify the effect of speech style and embodiment.

Takeover performance. Takeover performance was further divided into takeover time and quality (Table 3.2-2). Takeover time was defined as the time interval between the issue of TOR and the automation deactivation (Dogan et al., 2017; Vogelpohl et al., 2018), either by using the toggle or pressing the brake. Takeover quality metrics examined in the present study were speed-related measures (maximum, minimum, and average speed), maximum lateral acceleration, and standard deviation of lane position (SDLP; excluding the construction event, which required lance changing). Smaller values in speed, acceleration, or SDLP indicate smoother and safer takeover reactions. All takeover quality metrics were calculated during the manual control period between the time participants initiated the manual control and the time when they exited the takeover zone, which was a fixed point marked in the scenario and did not depend on participants' maneuver variation.

Each participant experienced four takeover events in each of the four scenarios, resulting in a total of 384 data points for each takeover performance measure. Values exceeding six standard deviations for each measure were revisited and corrected if a programming error was detected or removed if a true outlier was determined. No more than 3% of the total number of data points were excluded from each measure, with the maximum lateral acceleration having the largest number of points removed (n = 11) – mainly because of the simulation running error. Because the construction takeover event required a lane-changing maneuver, measures for this event were analyzed separately. The measures for the other three events were integrated and analyzed together. A two-way repeated ANOVA was conducted to determine the effect of speech style and embodiment on each measure for each event category.

Table 3.2-2 Takeover performance measures.

Category	Dependent Measures	Unit	Definition
Temporal measures	Takeover time	Seconds	Time between TOR and automation deactivation.
Takeover quality*	Max/Min/Average speed	m/s	Maximum, minimum, or average speed during the takeover event after automation deactivation.
Takeover quality	Maximum lateral acceleration	m/s^2	Maximum lateral acceleration during the takeover event after automation deactivation.
Takeover quality	SDLP	Meters	Standard deviation of the lateral distance of the ego vehicle regarding the middle of the lane.

* All takeover quality was calculated within the manual driving time frame within the takeover event zone defined along the route.

3.2.1.5 Procedure

Upon arrival at the lab space, participants signed the consent form for this study approved by the university's Institutional Review Board. Participants were explained that there were four driving scenarios with conditional automation where they would not operate the vehicle for most of the time. However, if the system asked them to do so, they had to take over the control and drive for some time before handing over the control back to the vehicle when prompted. To simulate a natural driving situation, participants were allowed to do any tasks of their choice during the drives, but they must be ready to take over the control when asked. Participants were also informed of the presence of the IVIAs, who would issue TORs and provide information regarding other road events. Before the formal drives, a simulation sickness test was administered (Gable & Walker, 2013), where a self-comfort checklist was completed before and after the 5-minute test drive. During this process, participants experienced sample

takeover events and a pause for SAGAT with sample queries, while also familiarizing themselves with the system control and simulated scenarios. Then, demographical information was collected if they were not suspected of simulation sickness. Before the first drive, the eye-tracking glasses were put on participants and appropriately calibrated. During the drive, the experimenter paused the scenario at two certain points—differed for each driving scenario—and administrated the SAGAT queries. After finishing each drive, participants completed the subjective questionnaires that collected their driver-agent interaction experience. After completing all conditions, participants ranked their preference toward four types of IVIAs and their reasons for the first and least preferred agents. The experiment lasted approximately 90 minutes.

3.2.2 Results

3.2.2.1 Driver-Agent Interaction

Table 3.2-3 summarizes the average rating score for each questionnaire under each condition. Embodiment showed a significant main effect on the System Response Accuracy scale: $F(1, 23) = 5.51$, $p < .05$, $\eta_p^2 = .19$, and in Trust in Automation scale: $F(1, 23) = 4.35$, $p < .05$, $\eta_p^2 = .16$. Robot agents were perceived to have higher system response accuracy and trust than voice-only agents, regardless of their speech style. No interaction effect between speech style and embodiment was identified in any factors.

Table 3.2-3 Subjective ratings on driver-agent interaction.

Scale	Items/Factors	Agent Type				Main Effect	
		IVA	CVA	IRA	CRA	Speech Style	Embodiment
SASSI	System Response Accuracy	4.96	4.82	5.28	5.17	-	V < R *
	Likability	4.85	4.79	4.74	5.07	-	-
	Cognitive Demand	3.18	3.03	2.98	3.13	-	-
	Annoyance	3.33	3.10	3.43	3.16	-	-
Trust in Automation	Trust	5.14	5.00	5.36	5.21	-	V < R *

* $p < .05$, I = informative, C = conversational, V = voice-only, R = robot.

3.2.2.2 Agent preference

Table 3.2-4 presents the preference ranking and distribution for each type of agent. A significant difference in participants' 1st preferred agent was found: $\chi^2(3) = 8.33$, $p < .05$. The conversational voice

agent was preferred the most. When participants explained the reasons behind their first preferred agent, explanatory or instructive (N = 7), less distracting (N = 5), friendly (N = 4), and human-like (N = 3) were the most frequently mentioned perceived characteristics of the conversational voice agents:

> *"The conversational voice agent provides the right information needed without being distracting as it may be with the robot or condescending with the strictly informative voice agent." (P5)*

> *"… remain relaxed and focused when having a voice that spoke in a friendly manner." (P10)*

> *"Being conversational engages the driver and provides correct instructions to prevent any human errors." (P12)*

> *"It is better to drive when the voice is more human-like and less of a robot. Also, the conversation makes it more lifelike as well. And explains more of what you should do." (P16)*

Table 3.2-4 Preference ranking for all agent conditions.

Preference	Agent Type			
	IVA	CVA	IRA	CRA
1st	6	**11**	**1**	6
2nd	6	6	6	6
3rd	5	5	8	6
4th	7	2	9	6
Average Score	2.54	1.89	3.04	2.50
SD	1.18	1.02	0.91	1.14

Note: unit – number of participants

When participants were asked to reason their least preferred agents, impolite/commanding/robotic (N = 7) and lack of information (N = 6) were raised frequently for informative agents, while distracting (N = 4) and uncomfortable (N = 3) were frequently mentioned for robot agents:

> *"The voice was very **robotic** and **unpleasant** to listen to. Also, the robot sometimes stares at me, which was **distracting**." (P9, 4th choice: IRA)*

> *"… the informative option was less ideal because it simply stated things that were obvious from the surroundings already." (P12, 4th choice: IRA)*

*"I feel like the informative voice agent was telling me what to do and **not in a polite way**. It felt like a backseat driver, which can get annoying." (P20, 4th choice: IVA)*

3.2.2.3 Situation awareness (SA)

A one-way ANOVA found that participants' transformed SAGAT scores differed significantly across scenarios: $F(3, 69) = 10.44$, $p < .001$. Thus, transformed SAGAT scores were further converted to Z-scores for each scenario before a two-way repeated ANOVA to understand the influence of speech style and embodiment.

Significant main effects of speech style [$F(1, 23) = 7.52$, $p < .05$, $\eta_p^2 = .25$] and embodiment [$F(1, 23) = 5.05$, $p < .05$, $\eta_p^2 = .18$] were found on participants' standardized SAGAT scores. Participants had a higher situation awareness score when accompanied by informative agents or robot agents than when accompanied by conversational agents or voice-only agents, respectively (Figure 3.2-2). There was no interaction effect between speech style and embodiment on participants' situation awareness.

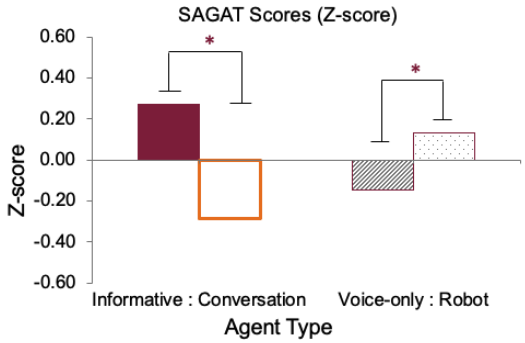

Figure 3.2-2 Standardized SAGAT scores across conditions. (p < .05)*

The influences of speech style and embodiment on each level of situation awareness were further evaluated through a 2 (speech style: informative vs. conversational) x 2 (embodiment: voice-only vs. robot) x 3 (SA: Level 1 - perception, Level 2 - comprehension, and Level 3 - projection) repeated measures ANOVA on the transformed SAGAT scores for each level. Results indicated that there was an interaction effect between speech style and situation awareness, $F(2, 22) = 5.96$, $p < .01$, $\eta_p^2 = .21$. Results from pairwise comparisons indicated that while participants were accompanied by conversational agents,

their Level 1 SA was significantly better than their Level 2 SA ($p < .05$) or Level 3 SA ($p < .05$); no difference was found between Level 2 SA and Level 3 SA ($p = 1.00$). Participants had higher Level 2 SA ($p < .05$) and Level 3 SA ($p < .01$) when accompanied by informative agents compared to conversational agents, while their Level 1 SA was not significantly different between speech styles ($p = .62$). No interaction was found between embodiment and situation awareness, $F(2, 22) = 2.53$, $p = .78$.

3.2.2.4 Eye-tracking measures

Results from two-way repeated-measures ANOVA indicated a significant main effect of embodiment on distraction fixation frequency (Figure 3.2-3): $F(1, 23) = 7.71$, $p < .05$, $\eta_p^2 = .25$. However, there was no difference in total distraction fixation duration between voice-only agent conditions (*Mean* = 293.37 sec, *SD* = 45.20 sec) and robot agent conditions (*Mean* = 290.69 sec, *SD* = 61.41 sec). Speech style did not influence the distraction fixation measures. When accompanied by robot agents, participants made more frequent distracting glances (*Mean* = 0.12, *SD* = 0.11) compared to when accompanied by voice-only agents (*Mean* = 0.07, *SD* = 0.06), regardless of the speech style, but the total distraction duration remained similar.

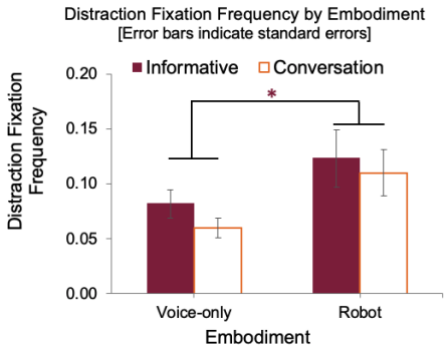

Figure 3.2-3 Distraction fixation frequency among all conditions. (p < .05) [Error bars represent standard errors]*

3.2.2.5 Takeover performance

Participants did not differentiate in their takeover methods (using a toggle attached to the steering wheel or pressing the brake) across agent conditions: $\chi^2(3) = 2.30$, $p = .51$. Thus, the subsequent takeover performance analysis did not separate these two methods.

Takeover time. Speech style or embodiment did not have a significant main effect on the takeover reaction time. The average takeover time was 1.46 s (*SD* = 0.28s) and 1.40 s (*SD* = 0.21 s) for the informative voice agent and the conversational voice agent, respectively; and was 1.42 s (*SD* = 0.21 s) and 1.42 s (*SD* = 0.22 s) for the informative robot agent and the conversational robot agent, respectively. Participants had a similar takeover reaction time across all conditions.

Takeover quality. For the construction takeover event, there was a significant interaction effect between speech style and embodiment on the maximum speed: $F(1, 23) = 5.32$, $p < .05$, $\eta_p^2 = .19$, and average speed: $F(1, 23) = 5.49$, $p < .05$, $\eta_p^2 = .19$. A simple main effect analysis indicated that when accompanied by voice-only agents, participants had a numerically higher maximum speed ($p = .067$), a significantly higher minimum speed ($p < .05$), and a significantly higher average speed ($p < .05$) when the agent communicated in an informative style compared to a conversational style (Figure 3.2-4).

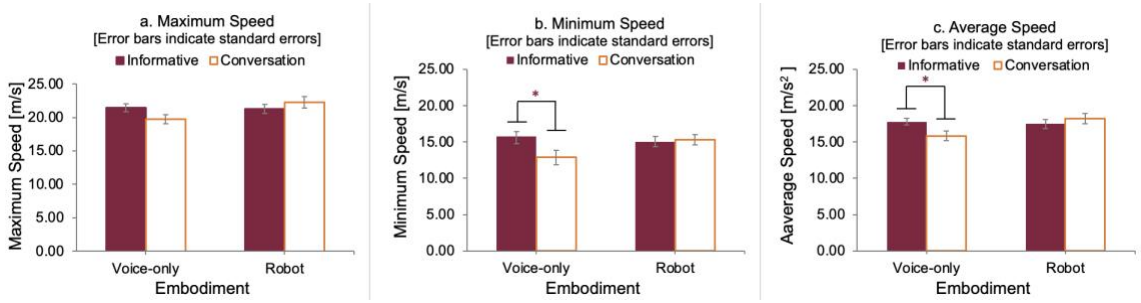

Figure 3.2-4 Max/Min/Average speed after takeover for construction. (p < .05) [Error bars indicate standard errors]*

Embodiment was found to have a significant main effect on the maximum lateral acceleration, $F(1, 21) = 4.36$, $p < .05$, $\eta_p^2 = .17$. Participants had a larger maximum lateral acceleration under robot conditions (Figure 5a). Speech style had a significant main effect on the standard deviation of lane position (SDLP), $F(1, 22) = 6.32$, $p < .05$, $\eta_p^2 = .22$. When accompanied by informative agents, participants had a higher SDLP (Figure 3.2-5b).

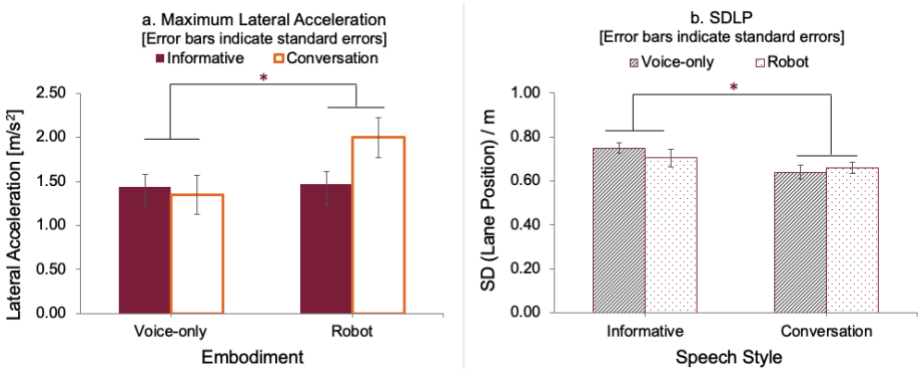

Figure 3.2-5 Max lateral acceleration and SDLP after takeover for construction. (p < .05) [Error bars indicate standard errors]*

In the non-lane changing takeover events (i.e., jaywalker, fog, and tunnel), speech style or embodiment did not have any significant main effect on the speed-related measures or SDLP. However, speech style showed a significant main effect on the maximum lateral acceleration, $F(1, 22) = 6.32$, $p < .05$, $\eta_p^2 = .22$. Participants had a larger maximum lateral acceleration under informative agent conditions (Figure 3.2-6).

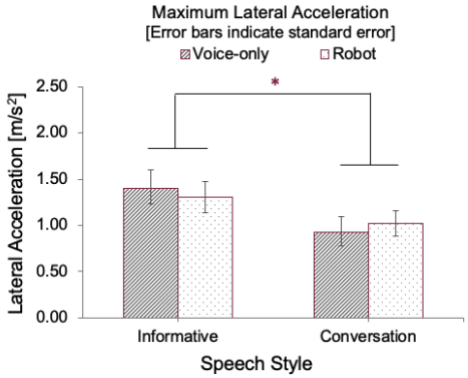

Figure 3.2-6 Max lateral acceleration after takeover for other events. (p < .05) [Error bars indicate standard errors]*

3.2.3 Discussion

This study investigated the effects of speech style and embodiment of in-vehicle intelligent agents (IVIAs) on drivers' experience and their takeover performance in a conditionally automated driving condition. Results indicate that although robot agents received higher ratings in driver-agent interaction questionnaires, they introduced extra distraction, which might cause suboptimal performance after the takeover. On the contrary, conversational agents gained driver preference and demonstrated consistent contribution to safer maneuvers after the takeover, compared to informative agents.

3.2.3.1 Speech style and embodiment on driver-agent interaction

While conversational speech style did not outperform informative one, the robot agents were rated more positively on subscales of system response accuracy and trust in automated systems. Although the information presented to drivers was the same between the voice-only and robot conditions, with perfect accuracy, the messages delivered by robot agents were perceived as more accurate and, therefore, met user expectations (Hone & Graham, 2000). Thus, higher trust in the embodied IVIAs was observed in our study as one of the benefits of elevated perceived accuracy that was used to form appropriate trust (Brink & Wellman, 2020). A previous study using the same robot found that NAO conditions overall yielded higher trust than voice-only agents, while a social NAO produced the highest trust score (Kraus et al., 2016). A review paper also pointed out that embodied agents are consistently more trusted (Rheu et al., 2020).

3.2.3.2 User preference favored the conversational voice agent

Findings from participants' preference toward four types of agents support our H1. Although participants rated robot agents higher in driver-agent interaction, they preferred voice-only agents the most (N = 17). Distraction and discomfort were two dominant perceptions when participants explained their preferences. The autonomous movement of NAO that made it "humanizing" (P4) was perceived as distracting due to its motion and noise. In addition, robot agents' occasional gaze focusing on participants made them uncomfortable. A long time being looked at by a robot counterpart may increase discomfort (Parreira & Gillet, 2022). Although the differences did not reach a traditional statistically significant level, subjective ratings on the subscale of discomfort (Table 3) were able to support these statements: robot agents were rated numerically higher on discomfort (IRA = 2.30, CRA = 1.99) than voice-only agents (IVA = 1.74, CVA = 1.85). When looking into the ratings on robot agents themselves, it seems that having a conversational speech style can mitigate such discomfort.

In fact, conversational speech style was also preferred the most (N = 17). Both the tone and contents provided by conversational agents were favored. As opposed to informative agents in a perceptually

"condescending" style, participants preferred the conversational agents speaking in a "friendly manner", which formed a feeling of being accompanied by "another passenger". In addition, participants preferred the "additional conversational part" that could provide "more context to its (i.e., the vehicle's) decisions". Thus, we could carefully speculate that conversational agents in our study provided higher perceived system transparency by improving drivers' understanding of agents' intentions (Bhaskara et al., 2020; van de Merwe et al., 2022; Wright et al., 2017). However, there was a debate about whether the additional information was necessary. Four participants claimed it was unnecessary because drivers were not heavily engaged; thus, they only needed to know the information when human input was required.

Overall, the conversational voice agent (CVA) as the most preferred agent (N = 11) retains the advantages of conversational agents without the disadvantages of robot agents.

3.2.3.3 Situation awareness and gaze fixation

We observed an interesting influence of speech style and embodiment on drivers' situation awareness. Although the conversational style was preferred the most and believed to provide additional content, drivers received lower SAGAT scores when advised by conversational agents compared to when advised by informative agents. Our H2.1 was not supported. Further analysis indicated that participants did not differ in their Level 1 SA-Perception between two speech styles but differed in their Level 2 SA-Comprehension and Level 3 SA-Projection, with the informative style having higher SAGAT scores. This indicates that although drivers were able to perceive the elements in the environment, they might not be able to develop better high levels of SA (i.e., comprehension and projection) when accompanied by conversational agents. Such performance decrement in preferred conditions may result from overreliance (J. D. Lee & See, 2004; Parasuraman & Riley, 1997) and complacency (Parasuraman & Manzey, 2010). Drivers were comfortable around conversational agents and satisfied with the information provided. Thus, they might not allocate adequate attention and cognitive resources to further analyzing the scenarios but rather depend on the agents to deliver information.

Similarly, although robot agents were commented to be distracting and annoying, drivers performed better in SAGAT queries when accompanied by robot agents. A potential explanation is that drivers may be able to prevent task-unrelated distractions. Although a higher distraction fixation frequency was observed under robot agent conditions, the presence of a robot only introduced a minimum level of distraction because the overall distraction duration remained similar. Thus, such distraction was not detrimental enough to compete with the driver's monitoring tasks. In this way, drivers were still able to overcome this interference and took compensatory actions, for example, improving their vigilance to the current situation in this study.

Research has indicated a positive correlation between fixation-related eye-tracking measures (e.g., fixation rate, fixation count) and direct SA measures (T. Zhang et al., 2020). Typically, Level 1 SA requires visual perception and can be inferred by focused and/or distributed attention. Such attention allocation can be assessed via eye-tracking measures (Jeon et al., 2024). In this study, both participants' distraction fixation frequency and distraction duration did not differ between speech styles, which aligned with the SAGAT scores in which their Level 1 SA scores were not significantly different. In terms of embodiment, although higher distraction fixation frequency was found in robot agents, participants' Level 1 SA scores did not differ between embodiment conditions, which aligned with the results related to distraction duration. This finding might suggest different capabilities of eye-tracking measures in predicting driver SA, as discussed in previous literature (Kim et al., 2020; T. Zhang et al., 2020). However, this study was not designed to explore the relationship between eye-tracking measures and direct SA measures. Thus, predicting direct SA measures using eye-tracking metrics was not the focus.

3.2.3.4 Conversational speech style produced careful maneuver

The takeover reaction time did not differ across conditions, which was not surprising because the TORs were delivered using the same set of messages. However, the takeover quality for the lane-changing event (i.e., construction) and non-lane-changing events (i.e., jaywalker, fog, and tunnel) was impacted primarily by speech style. At the same time, the embodiment also played a role independently of and dependently

on speech style. In general, drivers advised by informative agents exhibit risky and unstable driving behaviors in terms of higher speed across all speed-related measures, a larger standard deviation of lane position in the lane-changing takeover event, and higher maximum lateral acceleration in non-lane-changing takeover events. Only the effects on speed-related measures were moderated by embodiment; the influence of speech style diminished when the agent possessed a physical body. When taking participants' comments into account, the informative speech style was, in general, "annoying" and "irritating", which may create angry drivers who tend to drive faster (Jeon et al., 2015; Steinhauser et al., 2018). Further, the discomfort from the presence of the physical body was so strong that it might weaken or even override the effects of speech style; for instance, participants had a larger maximum lateral acceleration when accompanied by robot agents than voice-only agents.

When taking all takeover quality measures as a big picture, conversational speech style overall yielded greater takeover quality, which supports our H2.2. Although embodied agents were more favored in subjective ratings, speech style was more decisive and powerful in encouraging safer post-takeover driver intervention.

3.2.3.5 Implications, limitations, and future work

Findings from the present study are able to provide guidance on designing IVIAs for conditional AVs to deliver road information and issue TORs, which provide evidence on differentiating needs and requirements of IVIAs in vehicles with different levels of automation (Wang, Hock, et al., 2022). The balance between user preference and overreliance needs to be considered to maximize user acceptance while minimizing system misuse. Additional information explaining the system's current action is also critical to building explanatory and transparent IVIAs, which are helpful in forming appropriate mental models (Kraus et al., 2015). In contrast to embodied IVIAs preferred in full AVs (Large et al., 2019; Wang, Lee, et al., 2021), robot agents lost their likability in conditional AVs, where drivers are still required to take action. Interaction and companionship are no longer necessary in the form of a physical body but can be sufficient in a polite and friendly speaking style without visual distractions. Additionally,

delivering information in a natural, equivalent, and easy-going way, such as the conversational speech style, can promote user experience and elicit empathy in drivers, and further lead to cautious driving behaviors. However, if a conversational style is selected to prioritize user acceptance and experience, the contents included in the conversation should be carefully drafted to avoid complacency. Future research is needed to manipulate content richness and identify the balance between user acceptance and performance to present IVIAs for conditional AVs.

Even though our study has provided valuable findings leading to promising implications, we acknowledge some limitations worth further exploration. First, 50% of the speech prompts were TORs and were delivered using the same prompts. While we controlled the effect of prompt lengths on takeover reaction times, such a setting may compromise the differentiation between the informative and conversational agents, leading to equivalent momentary reactions to the speech prompt and similar user perceptions of two speech styles in driver-agent interactions. Second, although all driving scenarios were similar in terms of route and environments, elements on the road prior to the SAGAT freeze point differed to some extent. Such differences might introduce the scenario as a covariate when evaluating the SAGAT score. Even though we balanced the match between agent conditions and driving scenarios and standardized scores prior to further analysis, such a variety in the difficulty levels of SAGAT queries might impact user perception in an unforeseeable manner.

As intelligent agents gradually penetrate our daily lives, some of them have been applied to vehicles without any automation, where driver tasks dramatically differ from vehicles with conditional or full automation. As a consequence, IVIAs' responsibilities also shifted. Now that we have studies investigating the social attributes of IVIAs across different levels of automation, future attempts could be made to explore the variability in user perception and preference towards IVIAs across automation levels. In addition, we found a potential emotional reaction towards the IVIAs in the present study. The capabilities of IVIAs to elicit emotional states could be further evaluated and implied to mitigate performance decrements in emotionally impaired drivers (Dingus et al., 2016; Jeon, 2016).

CHAPTER 4 REQUIREMENTS GATHERING: EXPLAINABLE INFORMATION SYSTEMS IN AUTOMATED VEHICLES

4.1 Study 3: Identifying Information Needs in AVs – Contextual Inquiries

A variety of design techniques have been consolidated to explore human-automation interaction design (Pettersson & Ju, 2017). To date, a majority of research on designing and evaluating in-vehicle interactive systems has utilized the driving simulator studies to understand how drivers use and are influenced by those systems. Qualitative methods, which are powerful in revealing user needs and design requirements, are typically overlooked. Focus groups and interviews are the two mostly used qualitative methods to understand different topics in the field of automated vehicles. Although these two methods are capable of providing valuable insights to support in-vehicle interface design, the acquired information can be limited because participants are not situated in the task. In fact, field study methods that can obtain in-situ experiences have not been utilized in general. Despite the fact that there are some naturalistic driving studies and field operational tests, the purpose of them is primarily to investigate driving behaviors and performance. Due to the special characteristics of vehicles and the safety concerns of the driving tasks, the adaptation of the field study methods is challenging but not impossible. Researchers have successfully conducted field studies using various qualitative methods. Meschtscherjakov et al. (2011) conducted a contextual inquiry, an ethnographic study, and a cultural probing study to capture the car space and understand how drivers interact with the in-vehicle information systems during the ride. The General Motors User Experience Design Team adopted the Contextual Design Methodology delineated by Holtzblatt and Beyer (2017a) to inspire the automotive HMI designs (Gellatly et al., 2010). The team conducted contextual inquiries with 30 participants occupying luxury-type vehicles, followed by several interpretation sessions that produced sequence models, artifact models, affinity diagrams, and work models to interpret their results. Although field study methods have been primarily used to understand existing user tasks, they have also been successfully applied to futuristic technologies. Krome et al.

(2016) simulated fully autonomous driving using the experimenter as a driver, aiming to understand user experience in future commuting services.

Understanding contextual factors that are situated in the specific driving context is of great importance for the design of interactive in-vehicle interfaces that promote human-automation interaction. The objectives of Study 3 in my dissertation research are to (1) understand the in-situ driver behaviors when driving with features equivalent to Level 2 driving automation systems, (2) identify pitfalls in current driver-automation interaction, and (3) extract insights to direct the subsequent prototyping. To achieve these goals, Study 3 applies the Contextual Design methodology and carries out several contextual inquiries to understand the current challenges faced by drivers operating vehicles equipped with Level 2 driving automation systems. The information acquired from the contextual inquiries can provide valuable insights to inspire the design of explainable interfaces in conditionally automated vehicles. The specific research questions of interest are listed below:

- How do drivers utilize the driving automation systems?
- What are drivers' comfort levels in adopting automation systems in their vehicles?
- To what degree do drivers understand the limitations and capabilities of automation systems?
- How frequently do drivers engage in (what types of) non-driving related tasks (NDRTs) with the driving automation system active?
- What are other challenges in using driving automation systems?

4.1.1 Methods

4.1.1.1 Participants

Participant recruitment information was distributed through listservs and advertisement boards after the contextual inquiry was approved by the Virginia Tech's Institutional Review Board (IRB# 23-922). Potential participants filled out a screening questionnaire that collected their information regarding driving experiences and their vehicle information including the vehicle's make, model, year, and trim

level (i.e., pre-packages groups of features, e.g., basic, premium, platinum, limited). The perception of risk and frequency of risk behavior (Dingus et al., 2014) were also assessed to ensure the safety of the experimenters.

A total of 10 eligible participants (2 females) aged between 19 - 52 years old (*Mean* = 27.80, SD = 9.86) were recruited for the contextual inquiry. Participants had an average of 9.20 years of driving experience (*SD* = 9.75). All of them owned and regularly drove a vehicle equipped with Level 2 driving automation systems, except one participant who owned a vehicle (i.e., 2023 Mazda CX-30) equipped with Level 1 driving automation system that was admitted accidentally. I still kept this person's data as the information collected was still relevant to the goal of this dissertation. Table 4.1-1 summarizes the specifications for the vehicles recruited in this study.

Table 4.1-1 Specifications of vehicles recruited in this study.

Manufacturer	Model	Year	Features*	# of Vehicles Recruited
Tesla	Model 3	2019	Full-Self Driving package	2
Subaru	Legacy	2023	Adaptive Cruise Control, Lane Centering	1
Subaru	Crosstrek	2019	Adaptive Cruise Control, Lane Centering	1
Honda	Civic Touring	2022	Adaptive Cruise Control, Lane Centering	1
Nissan	Rouge	2020	Adaptive Cruise Control, Lane Centering	1
Hyundai	Sonata	2022	Adaptive Cruise Control, Lane Centering	1
Lexus	ES	2019	Adaptive Cruise Control, Lane Centering	1
BMW	X5	2022	Adaptive Cruise Control, Lane Centering	1
Mazda	CX-30	2023	Adaptive Cruise Control	1

*Different manufactures have different branding names for these features. Here I used the generic terms.

4.1.1.2 Procedure

The contextual inquiry consisted of an introduction session, a transition session, a contextual interview, and a final debriefing session, which lasted approximately 90 minutes. The description for each session is listed below. The experimenter script and detailed questions asked can be seen in Appendix B.

Introduction. The key to the success of the contextual inquiry is the comfort and naturalness perceived by the participant. Thus, participants went through a conventional interaction before the formal session, where they were introduced to the experimenters and the focus of study. Then, participants went through

the informed consent process to ensure that they understood the confidentiality. Participants' permissions were asked again for video- and audio-recordings. Participants were asked several questions on how the engagement with the driving automation systems fitted into their regular driving tasks through a brief semi-structured interview that lasted approximately 15-20 minutes.

Transition. After the short interview and before the actual driving trip, the participant was explained about the rules for the contextual interview (Holtzblatt & Beyer, 2017b). The participant was informed to drive on pre-selected routes where they would activate the driving automation systems while the experimenter actively observed their behaviors. The participant was informed that the experimenter will interrupt them whenever the experimenter identifies something interesting; but they could tell the experimenter if the timing was bad to be interrupted. The participant was also requested to think aloud every time they manipulated any feature in their vehicle.

Contextual Interview. After the rules of the contextual interview were clarified, the participant started driving. The experimenter prompted the participant with questions regarding the Level 2 driving automation systems, vehicle alerts, and their usage habits.

Debriefing. Any unanswered questions were asked during this session. I also asked several follow-up questions about the driving automation system if events of interest (e.g., system failures, activation of the automation systems) did not happen during the contextual inquires.

4.1.1.3 Data Analysis

All sessions were transcribed, and quality checked by two undergraduate research assistants. Based on the transcription participants' think-aloud responses and their responses to questions were extracted into notes that captured their opinions, attitudes, and descriptions related to their experiences with the ADAS. The generated notes were further transferred to sticky notes and used to create affinity diagrams to converge topics and themes that can foster the understanding of user needs and discussions related to building explainable intelligent agent prototypes.

4.1.2 Results

In total, 484 notes were generated, and labels were created for clusters with similar ideas or topics. In summary, there are 10 overarching categories merged from this affinity diagraming activity. Table 4.1-2 summarizes the topics under each category. It is noted that one note might be categorized under different affinity groups; thus, the total number of notes under all categories is greater than 484.

Table 4.1-2 Categories and Topics Merged from Contextual Inquiries.

Categories	**Topics**
User Description of ADAS	/
Overall Subjective Evaluations on ADAS	- user satisfaction - concerns and skepticism (e.g., dependence on features) - trust in ADAS (e.g., trust development, conditional trust, and distrust) - suggestions for improvement
System Usage and Frequency	Usage according to: - road type - temporal features of driving - traffic condition - driver state - weather condition
System Setting	/
System Limitations	- speed restrictions for system activation - weather restrictions - object detection restriction - road type and terrain restrictions - maneuver limitations - traffic condition restrictions
System Feedback and Display Components	- alert drivers about their unsafe behaviors - indicate the system's current status - indicate the system's understanding toward environment and other road users - indicate the system's understanding of human drivers - inform drivers about the system's availability - inform drivers about vehicle's future intentions - support navigation task
User Attitudes toward the Feedback and Display	Positive attitudes toward the following information: - prospective road information and vehicle action - vehicle perception - system and vehicle status - driving support Negative attitudes from the following aspects: - lack of information or explanation of certain vehicle behaviors - information overloading/receiving unnecessary information - annoyance - temporal characteristics (presenting alert too early or too late) - feedback design evaluations (e.g., not discriminable, inconsistent, etc.)
Learning and Training Process	Participants learn about their systems through the following approaches: - trial & error and past experiences - online resources - user manuals and update notices

Categories	Topics
	- learning curve - information from dealership is very limited
Past Experiences and Use Case Scenarios	- system failures - detection or recognition errors - interaction with other road users - feedback from systems - system driving performance (how well the system can keep the vehicle in the lane and how intuitive and smooth the adaptive cruise control is)
General Attitudes toward AVs	- positive attitudes toward AVs - distrust the capabilities of future AVs - potential technical limitations and areas for improvement

4.1.2.1 User description of ADAS

All participants were able to identify the ADAS technologies that were available to their vehicles. Some participants were able to use more formal names of the functions (e.g., "adaptive cruise control," "lane keeping assistance"), while the others might use more colloquial languages to describe the features (e.g., "keep you automatically away from the cars at a certain distance").

4.1.2.2 Overall subjective evaluations of ADAS

Although participants showed positive attitudes towards the driving automation systems, they also expressed concerns and wished the system could improve in several perspectives.

Participants expressed their satisfaction toward different aspects of the systems, including **well-visualized display** (e.g., P2: "They do a really good job of visualizing lots of things"), **reliability** (e.g., P3: "I've never had doubts about it," P13: "You don't really need to pay too much attention to small turns**"**), and **feature preference** (e.g., P14: **"**It was the thing I liked"; "it" here refers to the lane keeping assist function).

In the meantime, participants also expressed concerns and skepticism toward using ADAS. These concerns were derived from the mismatch between vehicle driving styles and drivers' driving styles, increased mental effort when trying to understand automation, glare from direct sunlight, etc. One of the frequently mentioned concerns was **dependence on features**. Five participants mentioned that they were

more used to the features that made them "lazy**,**" "more tired**,**" and more likely to "remove attention" from the driving task (e.g., P3: "It made me more prone to watch my phone").

There were also discussions regarding trust in automation systems. Five participants mentioned that their trust was developed gradually as they interacted more with the systems (e.g., P2: "I definitely have gotten much more trusting of it with greater experience"). Throughout those interactions, participants also developed **conditional trust**, meaning that they would choose to trust the systems under certain conditions (e.g., P17: "I trust it to drive in a straight line more than a super curvy road"), while they would prefer to have the control under other conditions where they have learned the systems might fail (e.g., P3: "so like sharp turns, I'll just hold on to it"). However, four participants also mentioned that they, in general, distrust the vehicle mainly due to these reasons: (1) preference to be the locus of control (e.g., P4: "I prefer my feelings of in control"), (2) system failures experienced or witnessed in the past (e.g., P9: "I haven't seen a system yet that I can trust enough to like"), and (3) unjustified distrust (e.g., P3: "I didn't trust the autopilot. I was like, very scared, I said, Oh my gosh. So, I didn't turn it on that much.").

Finally, participants also proposed suggestions to improve ADAS, including matching automation system driving styles with drivers' driving styles, improving system reliability (e.g., "fewer errors"), informing users about the system's status, and combining some features into one display element (e.g., combining maps and cruise control indicators).

4.1.2.3 System usage and frequency

Participants discussed several factors that might influence their decisions to engage in driving automation systems. These factors, ranked by frequency of mentioned, include **road type**, **temporal features** of driving, **traffic condition**, **driver state**, and **weather condition**.

Road type: while most participants stated that they used the adaptive cruise control function mostly on freeways, some participants kept the lane assistance always on while driving. Roads with "too many

stoplights," "sharp curves," and "traffic circles" might prevent participants from using either or both driving automation systems.

Temporal features impact drivers' decisions. Participants, in general, chose to use those functions during longer drives "over 3 miles", "any car rides like 30 minutes", or "long road trips." One participant also mentioned the time of day and preferred not to use the system during the morning, considering that "people aren't all the way awake."

Traffic condition refers to other road users sharing the road, which can be used to determine whether it is a "safe location" to activate the systems. Participants preferred not to use systems under traffic with "erratic vehicles," "pedestrians," or other unpredictable objects surrounding them. However, there were different preferences in using the systems under "stop and go traffic" or "when the traffic is worse": a few participants preferred to use the systems under such situations because they "trust the car a little bit more than myself," while others preferred not to use because they "would rather be in control."

Driver state refers to the driver's mental state. Tiredness and comfort with system operation are two factors identified in this category. One participant mentioned that they tended to use the features when they were **"**tired and don't feel like driving." Two participants mentioned that they would use the system when they felt comfortable doing so.

Weather conditions also play a role in drivers' decision-making. Participants would not use the system under severe weather conditions (e.g., "snowing," "super-bad" rain), while they would still use them under mild unclear weather conditions (P3: "I do this when it's raining, but I also like to pay a good amount of attention and always have my foot on the brake just in case.").

Participants' own **judgement on system operation** refers to whether they thought the system could function properly.

4.1.2.4 System settings

Most users who mentioned their system settings were able to clearly articulate how to set up the parameters for their driving automation systems (e.g., the distance between the following vehicle) or the in-vehicle interface setting (e.g., which interface to display). However, one participant kept explaining that everything was set to default, and he did not change too much or was not aware of any setting, although he had owned the vehicle for six months.

4.1.2.5 System limitations

Participants were asked whether they were aware of any limitations or restrictions that could potentially prevent them from activating the ADAS. The discussed limitations include **speed restrictions for system activation**, **weather restrictions**, **object detection restrictions**, **road type and terrain restrictions**, **maneuver limitations**, and **traffic condition restrictions**.

Speed restrictions for system activation. Four participants (two Tesla owners, one Subaru Legacy owner, and one Hyundai owner) stated that they did not believe there was a minimum speed requirement for their system activation. However, two participants mentioned that the target speed of adaptive cruise control could not be set under 20 mph. One participant stated that their vehicle's adaptive cruise control can only be set when the vehicle was traveling more than 50 mph. As for the lane centering assistance, participants stated that their systems could not track the lines until a certain speed was reached (e.g., 40 mph, 30 or 35 mph). One participant also reported that they were aware of a speed restriction but could not remember the exact number. One of the Tesla drivers also reported that 80 mph was used to be the maximum speed for the autopilot to work, but this parameter might be increased to "85 or 90" in recent updates.

Weather Restrictions. Six participants stated that their systems still worked well under light rain and light snow conditions, but one participant also stated that they "personally wouldn't do this when it is

snowing." Two participants also mentioned that their vehicles would disable the functions when the weather conditions did not allow clear camera vision.

Object Detection Restriction. Participants mentioned occasions where their systems had limited capability in object detection or recognition. For example, two participants mentioned that their systems were not able to trace the lines under certain conditions (e.g., faded lines, exit lanes). Two participants also mentioned that the systems were not able to accurately update the speed limits accordingly. According to participants, inaccurate mapping and the inability to detect pedestrians or crossing traffic were two types of objects that the systems were not able to detect.

Road Type and Terrain Restrictions. Participants reported system failures under different types of roads and terrain, including road segments with sharp curves, parking lots, and construction zones. Two participants also pointed out that the ideal scenarios for those systems should be highways, clear roads, clear weather, and no traffic signals.

Driving Maneuver Limitations. One Tesla driver mentioned that the system was not currently capable of making U-turns and was not very skillful in navigating through roundabouts.

Traffic Condition Restrictions. Two participants also discussed the system's capability in stop-and-go traffic, while one of them stated that their system worked fine but had a restriction on stopping time (P9: "This does turn off if the cruise brakes on its own under its own power for like 10 seconds.") and one of them did not think the system could work under such conditions.

However, four participants mentioned that there were no restrictions mentioned above that would restrict them from activating the systems.

4.1.2.6 System feedback and display components

Throughout the contextual inquiry, most of the conversations were centered on the system feedback and display components, which can be categorized into the following aspects, ordered in frequency: (1) alert

drivers about their unsafe behaviors, (2) indicate the system's current status, (3) indicate the system's understanding toward environment and other road users, (4) indicate the system's understanding of human drivers, (5) inform drivers about the system's availability, (6) inform drivers about the vehicle's future intentions, and (7) support drivers' secondary driving tasks.

Alert drivers about unsafe driving behaviors. The most frequently discussed system feedback is to alert drivers about unsafe behaviors that can potentially lead to dangerous outcomes. Departing from the current lane was the most common alert, which typically consisted of both visual and auditory warnings at the same time. If the vehicle also had the heads-up-display (HUD) available, such an alert would also be present on the HUD. Two participants also reported haptic feedback associated with their lane departure warning.

Indicate the system's current status. Participants discussed icons or other indicators that showed the current status of the driving automation system. Specifically, these indicators provided information regarding system on/off (e.g., whether the system is actively steering), system readiness (e.g., whether the adaptive cruise control meets the operation requirements and is ready to use), current state of the system among multiple states (e.g., for lane centering assistance on Subaru: grey indicates not tracking the lane, white indicates actively tracking the lane, and green indicates actively steering), and the current system setting (e.g., following distance). All of these indicators were presented using visual icons or elements on the HUDs if the system remained in its current state. However, if the system states changed, an auditory cue might be present for certain functions, such as the short beep indicating the steering assistance was no longer actively working.

Indicate the system's understanding of the environment and other road users. The most representative feedback that indicated the system's understanding of external stimuli was Tesla's visualization of vehicle surroundings. Other vehicles interviewed in this study, including the BMW X5 and Honda Civic Touring, also had similar visualization but were not as comprehensive (i.e., had limited

object recognition capability). These displays could update the elements in the environment in real-time. Except for the real-time updates, most vehicles were capable of tracking discrete states of external stimulus, including road sign changes, leading vehicles within/out of the tracking range, weather conditions (e.g., P16: “the rain or snow is too heavy, clear the camera before use.”), road condition (e.g., P16: “It was warning me that there was a big curve ahead.”), proximity detection (e.g., obstacles while traveling, obstacles along the way of opening the doors), and traffic light status. While most of these elements were presented as visual components on the vehicle instrumental panel or the HUDs, the change of external environmental status would also trigger auditory alerts with various perceived urgency. For instance, drivers received a non-urgent auditory cue regarding the leading vehicle entering and exiting the tracking range when the adaptive cruise control was on. If the vehicle detected a potential forward collision, an urgent auditory alert with a visual warning message would also be present in most vehicles. Some of the information was also available on the HUDs if the vehicle was equipped with one.

Indicate the system’s understanding of human behaviors. All vehicles recruited in this study were equipped with driver monitoring systems of various capabilities. The most common driver monitoring discussed was the hands-off-the-wheel warning, which took place in several stages incorporating different modalities. The time between the offset of hands and the initiation of the warning varied depending on the vehicle made. In general, the system first issued a visual warning message to ask drivers to put their hands back on the steering wheel. One vehicle equipped with light bars on the steering wheel would also flash yellow if no hands were detected for a certain time. If no action was taken, an auditory warning would be present. Systems on the Hyundai were the only ones that would also issue a haptic warning. Finally, two Tesla drivers and the Nissan driver also reported that the vehicle would slow down after not putting their hands back on the wheel for an extended time period. In addition, the BMW X5 was the only vehicle interviewed that was equipped with in-vehicle cameras to track drivers’ eye gaze when traffic jam assist was active (e.g., “When you're using it with traffic jam assist plus where it steers for you, you have to pay

attention constantly or else it kind of gets mad.”). The BMW X5 would also propose break times to drivers if the engine started for over three hours.

Inform drivers about the system’s availability. Participants also discussed the feedback they received when their driving automation systems were not available to use due to severe weather, road type restrictions, and lock-outs due to unsafe driving practices. This type of feedback was presented using both visual and auditory components. However, not all feedback was clearly explained, and drivers were still able to find rationales behind the system's unavailability.

Inform drivers about the vehicle’s future intentions. One of the Tesla drivers discussed several occasions where their vehicle would inform its future actions, which were presented using both text messages and elements on the visualized vehicle surroundings.

Support drivers’ navigation task. Two out of ten vehicles recruited could also integrate the navigation information within the instrumental panel or HUDs, supporting drivers’ secondary driving tasks directly compared to other vehicles where navigational displays were separate.

4.1.2.7 User attitudes toward the system feedback and display

Participants expressed both positive and negative subjective evaluations of the system feedback.

Participants expressed that presenting vehicle perception gave them confidence, which includes presenting vehicle perception towards both the external environment and internal driver states. Having the information regarding the system and vehicle status was also helpful in keeping participants updated. Participants also found the prospective road information and vehicle action helpful in making decisions and planning (see the direct quotes below). Participants’ general attitudes toward the driver monitoring system (i.e., alert drivers when hands off the wheel) were positive, although this type of alert could cause annoyance (P3: “I found it very annoying when it tells you to put my hand on it. But I realized that it's pretty good.”).

"Especially when you're about to go over railroad tracks, some cities that I drive and just have the railroad tracks built into either the highway or the road, and that's really helpful. Just make sure you don't, you know, speed Through that or anything." (P16 favored the knowledge about prospective road information)

Participants also complained about the feedback and displays of their vehicles from certain aspects. The most frequently mentioned aspect was the lack of information or explanation on certain vehicle behaviors. Typically, these vehicle behaviors did not match with users' own perception (e.g., the vehicle swerved to avoid an obstacle that was not visible to the user, or the signs shown in the vehicle display did not match with what the user saw on the road), which results in user confusion. Participants also mentioned that they sometimes experienced information overload (e.g., P16: "the last thing I want when I'm driving is to read and stuff") and received unnecessary information when they were in control (e.g., P17: "Because when you're driving, you want to be looking at the road and what's happening currently in front of you and not what the car sees."). In line with this, participants reported annoyance when receiving too many auditory alerts (e.g., P9: "I get annoyed about the pop-ups a little bit because it's just like I get it. I'm already going like, I'm fine."; P14: "If I'm not using the adaptive cruise control, it just makes noises."). The feedback timing was also reported to negatively impact user experience if presented too early or too late. Presenting too early might annoy users (e.g., "Some of the preemptive pop-ups are a little fast or a little ignorant to what you're actually doing."). Presenting too late can hurt user trust ("It is not trustworthy that sometimes you had latency with adjusting itself with the signs of the road."). Finally, participants evaluated the feedback from the perceptual perspective: auditory cues in some vehicles were not discriminable and thus not informative (e.g., P16: "Using the same sound for everything is not a great idea. The sound tells you that something's going on, but I need to use my screens to figure out what exactly that is."). The inconsistency of the icon locations could also confuse users.

In addition to positive and negative experiences, participants also proposed additional features that they would like to have. For instance, one participant suggested that the in-vehicle camera could also be used

to detect driver distraction and drowsiness rather than just determining driver eyes on the road when the traffic jam assistant was in use. The same participant also suggested that the visualization of their vehicle would be better if the road curvature were considered rather than just showing a straight road all the time (see the direct quote below). In addition, more information regarding the traffic signals was also preferred, such as the state of the stop light and additional feedback (e.g., haptic) when they were ready to proceed at the controlled intersections.

> *I just want to know if this car can see the corner that's up ahead and like, like, how steep it is because, again, it really sometimes feels like this car has no idea there's a corner, and the last second it just takes it. (P16, comments regarding vehicle surrounding visualization)*

4.1.2.8 Learning and training process

During the contextual inquiry, one of the main topics that we discussed was how participants learned about the driving automation features and how long it takes for them to get familiar with the system. Learning from **trial and error and past experience** was the most frequently mentioned learning pathway, followed by exploring **online resources** and reading **user manuals and system update notices**. Participants also mentioned they did not receive much information from the **dealership** when purchasing.

Trial & Error and Past Experience. All participants mentioned that they learned how the system works from trial and error or their past experience with similar vehicles. Participants "played around" the functions and "tried them out" when they first purchased the vehicle or when a new feature was released.

Online resources. Six out of ten participants mentioned that they also learned through online resources, such as online videos (including YouTube and other video sources), user forums, and direct information searches on Google.

Limited Dealership Support. Five out of ten participants mentioned that they did not receive any instructions or "features run down" through the dealership. One participant (P13) also reported his

negative experiences with the dealership: “What they're saying is like we're only selling the car; we don't really know anything.”

User Manuals/System Update Notices. Only four out of ten participants expressed that they went through the user manual. One participant—one of the Tesla owners—also read through the system updates.

In terms of learning time, participants reported that they learned some features in a really short time, about two to three long drives or “about a week.” However, some features or some specific aspects of a certain feature required a longer time to grasp, such as the autopilot and full self-driving in Tesla and the distance setting in the adaptive cruise control (e.g., P13: “It took me some time to figure out how to adjust to different distances”; this participant was talking about how to set up the following distance for the adaptive cruise control).

4.1.2.9 Past experiences and use case scenarios

Learning through trial and error and past experiences was the most frequently mentioned pathway to getting familiar with the systems. Participants’ statements related to their past experiences and some use case scenarios were further reviewed. I found that those experiences include **system failures**, **detection or recognition errors**, **interaction with other road users**, **feedback from systems**, and **system driving performance**.

System failures refer to situations where the system is no longer available**.** Participants reported several occasions where the systems might turn themselves off without clear explanations. Some of the examples include traveling over a certain speed that the system turned itself off without explanations, foggy areas where the camera systems failed, and the system not working well under certain road types (e.g., roads without clear markings, sharp curves with high speed), and construction zones that the system was overwhelmed.

Detection or recognition errors refer to when the system was still working but made mistakes in detecting or recognizing objects on the road. These errors include (1) false alarms where the systems detected certain objects on the road, but the human drivers did not detect them; (2) misses where the systems were not able to detect objects on the road due to terrain restrictions (e.g., sharp curves, going up the hill), weather conditions (e.g., direct glare from the sun), or other vehicle types (e.g., a truck with a trailer); and (3) inaccurate road sign recognition.

Interaction with Other Road Users. One of the Tesla drivers discussed a lot about how their vehicle interacted with pedestrians. This participant mainly discussed that the vehicle had a difficult time making decisions when encountering pedestrians, which led them to turn off the automated driving systems.

Feedback from Systems. Participants shared occasions where they learned about certain system limitations or how the system communicated their intentions and supported drivers. For instance, one participant discovered that their system would disable the whole driving automation system under severe weather conditions. Another participant articulated how their system updated itself in terms of visualizing vehicle surroundings and providing feedback to explain vehicle intentions, with an example of the vehicle creeping forward to make a safe right turn.

System driving performance refers to the automation system driving performance (i.e., the smoothness of automated operation). Throughout repeated experiences, participants also learned how their automated driving systems performed under certain conditions. Such performance includes lane-changing behaviors (e.g., P2: “It does not like to change back into the slow lane, so it will just stay in the fast lane unless…”) and system boundaries (e.g., P9: “If you're getting really close to the line, it will pop in the message”).

4.1.2.10 General attitudes toward automated vehicles

When asked about the general attitudes towards AVs, participants hold diverging opinions. Half of the participants (5/10) expressed positive attitudes toward AVs. Participants were “open,” “excited,” and “positive” about future AVs and thought they were “cool.” Although positive, two of them also expressed

some concerns: one of them was “not sold on the fully automated like hands off the wheel,” while the other participant claimed that AVs would do better in simultaneous object recognition but might not do better than humans in other tasks.

Three participants explicitly expressed their distrust towards AVs (e.g., “I cannot trust them,” “It would be cool, but I just don’t fully trust it,” “It cannot update itself as a person can, or as vigilant as a person can”). One participant also mentioned their concerns regarding people’s overreliance on technology, “I'm worried that, you know, as we see more of these features, we're gonna see more people doing that.”

One participant also expressed their proposals for potential technology improvement. Informing users about “what it (i.e., the AV) is doing” and considering driving styles in system design.

4.1.3 Discussion

Throughout the contextual inquiries, I have developed a better understanding of how drivers currently utilize the driving automation system in terms of when and where. Although most participants adopted their ADAS on highways where the systems were designed to be used, some drivers still used them in places outside of the ODD of the ADAS. The systems do not put any constraints on system activation in terms of road type but only on severe weather, which aligns with the findings shown in Monticello (2023). This finding also indicated that most participants were not fully aware of the limitations of the ADAS technologies and thus engaged in potentially unsafe behaviors either by activating the ADAS where they are not intended to be used or engaging in NDRTs more often. Such user behaviors are also aligned with previous findings that drivers often develop inaccurate mental models of ADAS 2/11/2025 3:04:00 PM and further lead to misuse or disuse of these systems. Even though participants in this study were aware of the risks associated with ADAS that lead them to be more dependent on the features, they still voluntarily chose to engage in unsafe behaviors.

Findings from the contextual inquiries also indicated that drivers learned about the ADAS technologies from a variety of sources, including user manual and other original equipment manufacturer (OEM)

produced materials (e.g., website and tutorial videos), trial and error, online resources, family and friends, etc. (Mason et al., 2023). However, all participants reported a lack of support from the dealership. Almost all participants preferred to learn about the systems through on-road trial and error, while some of them also went through user manuals. We did not observe any impact of age differences on drivers' preferred approach to learning as indicated in Abraham et al. (2018) but rather aligned with the finding in Eby et al. (2018), where older adults were also found to prefer trial and error.

In terms of trust, most of the participants were positive about ADAS and future AVs. Through trial and error, participants developed justified trust and distrust towards system capabilities, especially after they had owned the vehicle for a longer time (Lubkowski et al., 2021). However, we also identified participants who maintained low trust in the system and general AVs due to unexpected ADAS behaviors (Lubkowski et al., 2021).

The primary challenge of improving the ADAS experience and fostering appropriate trust and adoption towards the driving automation systems lies in how to effectively design in-vehicle interfaces to communicate vehicle limitations while balancing the information load and information value provided. Large individual differences were also identified in terms of preference for receiving system feedback, which also needs to be considered when designing systems to address user needs.

Although insightful, this study also has a limitation in terms of sampling bias. To ensure safety, only drivers who did not exhibit risky driving behaviors in the pre-screening survey were recruited in this study. We also required participants to regularly use their ADAS so that they could articulate during the contextual inquiries. Therefore, participants recruited in this study were more likely to be safe and well-educated drivers (in terms of ADAS) and do not necessarily represent the general driver population who might carry out more frequent unsafe driving behaviors. However, even with safe drivers, we were still able to identify issues related to driving automation systems and potential opportunities for improvement.

4.2 Study 4: Solutions to Explainable Interfaces from Expert Interviews

Experts are defined as "people who possess special knowledge of a social phenomenon which the interviewer is interested in" (Gläser & Laudel, 2009). The expert interview is a form of interview—usually in the form of a semi-structured interview (Meuser & Nagel, 2009)—targeting people in certain professional positions from whom information about professional processes can be gathered (Flick, 2023). There are three types of expert interviews: **exploratory**, **systematizing**, and **theory-generating** (Bogner & Menz, 2009).

The exploratory expert interview is helpful to establish orientation in fields under investigation and to generate hypotheses. The purpose of exploratory expert interviews is not to acquire as much information as possible or compare data, which distinguishes this type of expert interview from the other two types (Bogner & Menz, 2009). For instance, Beringhoff et al. (2022) conducted a semi-structured interview with 13 experts from industry and academia to understand the methods and tools used to test automated vehicles (AVs), where 31 challenges were identified, with 26 of them not known previously. Their findings indicated that although some of the challenges have been discussed in existing research efforts, challenges in the field of scenario-based testing and simulation were still unresolved at the time of publication (Beringhoff et al., 2022).

The systematizing expert interview—the most widespread form of expert interviews—aims to gain access to knowledge exclusively possessed by experts and to obtain systematic and complete information (Bogner & Menz, 2009). It is critical for the systematizing expert interviews that the data collected are comparable to the subjects of researchers' interest. Tabone et al. (2021) conducted a relatively structured interview to systematically gather information from 16 Human Factors research on the topic of interaction between AVs and vulnerable road users. They structured the interviews from four themes: general questions on AVs, external human-machine interfaces (eHMIs), augmented reality (AR) and AR eHMIs, and virtual reality (VR) and AR experiments, and compared results from these topics across interviewees (Tabone et al., 2021).

Finally, the theory-generating expert interview aims at developing typology or a theory about the topic that researchers are interested in through analytically reconstructing the knowledge from multiple experts (Bogner & Menz, 2009; Flick, 2023). Lee et al. (2022) proposed a draft design taxonomy of user needs in future AVs and finalized the taxonomy through expert interviews with nine experts and focus group interviews with ten attendees.

The expert interviews carried out in this dissertation primarily aimed to address the solutions for the given challenges. Thus, the systematizing expert interview is the most suitable interview type. The following sections delineate the interview structure and data analysis plan.

4.2.1 Methods

4.2.1.1 Experts Qualifications

Explainable artificial intelligence (XAI) is a human-agent interaction problem that intersects across subjects of Social Science, Artificial Intelligence (AI), and Human-Computer Interaction (HCI) (Miller, 2019). Thus, experts from these three domains with relevant experiences with XAI and AVs will be identified and recruited for the expert interview. Researchers from industry or academia who actively engage in research on explainable interfaces for automated vehicles will be considered experts if they meet the criteria listed in Table 4.2-1.

Table 4.2-1 Criteria for selecting experts as interviewees.

Experts in Academia	Experts in Industry
• Holds a doctoral degree OR Is a research scientist at a research-oriented university or at a research institute with at least five years of research experiences. • Actively conducts research in related fields. • Research involves developing and investigating explainable interfaces for AVs.	• Holds a doctoral degree OR Has at least five years of experience in automobile industry. • Actively conducts testing related to in-vehicle user interfaces. • Testing involves novel interfaces for advanced driver assistant systems.

4.2.1.2 Expert Demographics

Experts were reached out after the interview got approval from Virginia Tech's Institutional Review Board (IRB#23-1216). Experts' email confirming their willingness to participate in this study was considered as their written consent. A total of seven experts were recruited, two of which were in the field of Social Science, one of them was in the field of AI, four of them were in the field of HCI (including two industrial experts). The experts included in my dissertation have an average of 16.71 years of experience in their field (*SD* = 13.28). Table 4.2-2 summarizes the research focuses of the invited experts and their research experiences relevant with explainability.

Table 4.2-2 Experts' experiences in the field and their field of studies.

Expert	Research Focus	Experience [Years]	Research Relevance with promoting explainability
1	Cognitive systems engineering, AI	45	Automation systems' workings and user comprehension; not specific to AI in automobiles but applicable.
2	Human-Machine Autonomy Trust in Automation Transparency in Human-AI Teaming	19	Importance of transparency in systems like command-and-control and space systems; operators need to understand machines' actions and assumptions.
3	Multitasking in safety-critical systems Pure explainability to practical implications in AI systems.	8	Past work: driving explainability (performance and trust through interfaces, e.g., AR) Current focus: subtle cues for understanding AI in continuous systems.
4	Human-machine interface (HMI) Driver performance Safety in automation systems	11	
5	Human Factors influences in automated driving; development of new interfaces for vehicles, including AI and passenger drones.	17.5	Technical side of automated systems User interaction with AV technologies.
6	AI-based decision support systems Machine learning and reinforcement learning techniques.	7.5	Utilizing explainable AI methods to design and evaluate solutions No experience in automated vehicles.
7	Driver performance related to Level 2 driver assistance features and collision avoidance. Driver supervision of semi-autonomous features.	9	Testing iconography comprehension for semi-autonomous systems Developing educational materials for better user interaction with these features.

4.2.1.3 Interview Procedure

To ensure each expert received the equivalent information in terms of the topic covered and terminology used, the interviews were carried out in a semi-structured form with the five segments described below. Experts were allowed to choose between interview formats: live interview or asynchronized interview (i.e., by filling out the interview questions and allowing follow-up questions). The latter format was provided to accommodate the experts' schedule or to accommodate the industrial experts whose participation might be limited due to company policies.

Each interview consisted of five sessions: introduction, general understanding of explanations for automation systems, discussion on specific user challenges, trends in XAI, and closing. The interview lasted for 60 minutes. The brief description of each section is listed below. The complete semi-structed interview questions can be found in Appendix C.

Introduction. The interview started with the introduction to the interviewers' background and the purpose of the interview, followed by several brief questions that survey the expert backgrounds.

General Understanding of Explanations for Automation Systems. This session explored experts' understanding of XAI, especially towards AVs if they have relevant experiences.

Discussion on Specific User Challenges. Expert read through two use case scenarios and expressed their understanding of the situation and their opinions to address the challenges.

Trends in XAI. Experts' opinions on future trend of explanations and their risks were collected.

Closing. After the interviewer covered all the discussion points, experts were given opportunities to express any ideas that were not covered.

4.2.1.4 Data Analysis

Information gathered from expert interviews was analyzed using the inductive thematic analysis (TA) method, which is "a method for systematically identifying, organizing, and offering insight into patterns of meaning (themes) across a data set" (V. Braun & Clarke, 2012, p. 57). The expert interviews were audio recorded and transcribed later to ensure all spoken words and sounds are captured, including hesitations, false starts, etc. Then, a six-phase approach was used to conduct the thematic analysis (V. Braun & Clarke, 2006).

Table 4.2-3 Six Phases to Thematic Analysis. [Adapted from (V. Braun & Clarke, 2006)]

Phase	Description
1. Data Familiarization	All the interviews were transcribed. The transcription was broken down into utterances based on the punctuation before roughly reviewing to come up with initial ideas.
2. Initial Codes Generalization	Based on each utterance, two coders generated initial codes using the original words from the experts as much as possible. These codes labeled the features of the data that were potentially relevant to research questions.
3. Themes Searching	After all interviews were coded, I searched for themes that captured "something important about the data in relation to the research question and representing some level of patterned response or meaning within the data set" (V. Braun & Clarke, 2006, p. 82).
4. Themes Reviewing	After extracting themes, I checked if the these were related to the codes identified (Level 1) and the entire interview datasets (Level 2), which resulted in a thematic "map".
5. Themes Defining and Naming	I continued to refine the themes to ensure a holistic view from all interviews. Definitions and names for themes were finalized in this phase.
6. Report Production	The thematic analysis results were reported in this dissertation in a structured manner, with exemplar quotes from experts, and the frequency of the themes mentioned by them.

4.2.2 Results

A total of 1763 utterances were produced from the structured expert interviews from seven experts. In total, there were seven major themes identified. The definitions of each theme and example codes associated with them is presented in Table 4.2-4. The following sections explained each theme and described any subthemes identified under the major themes, if any.

Table 4.2-4 Identified themes, definitions, and example codes.

Theme	Definition	Example Codes
Consequences of badly designed human-machine interface	Uncover existing unintended consequences of current advanced technologies.	break neck to see; confusion
Explanation mechanisms and adaptability	Highlight the role of explanation mechanisms in supporting user understanding and user perception, and the need for these mechanisms to adapt to changing user needs and contexts (i.e., what explanation is and should be).	instantaneous learning curve;
Methodologies, models, and frameworks for in-vehicle explanation design	Discuss the methodologies that are already available in other field and can be used for in-vehicle explanation design.	law of cognitive systems;
Traps and tricks in in-vehicle explanations design	Emphasize some "rabbit holes" or precautions that involved in the in-vehicle explanation design. Focusing on what the explanations should not be or some traps that the researchers might fall in when designing explanations.	no interface developing;
Design considerations for in-vehicle explanation design	Focus on how to make proper design decisions for explanations in AVs, describing guidelines or recommendations for in-vehicle explanation design.	tradeoffs; contextual factors
Risks associated with adding explanations	Discuss the potential risks after adding explanations into the current setting.	manufacturer fault; forced explanations
Educating and training AV users	Address the need for ongoing education and training for users to adapt to advanced automated vehicles and maintain necessary skills.	training; interactive tutorials

4.2.2.1 Theme 1: Consequences of badly designed human-machine interface

Six out of seven experts shared their experiences or opinions with badly designed human-machine interfaces under the context of automation systems that were not necessarily limited to automated vehicles. This major theme consists of three subthemes: (1) negative experiences with badly designed interfaces, (2) user confusion, and (3) decreased user acceptance and trust.

Negative experiences with badly designed interfaces. Experts described their own and their family's or friends' negative experiences with badly designed interfaces. Such experiences included suboptimal interface design that required additional user effort to take actions, hidden switches for the automatic engine shutdown function and automatic high beam adjustment, driving automation systems turning themselves off without further explanations, automatic emergency brakes without considering the

following vehicle distance, etc. All of these negative experiences can contribute to user confusion and further decrease user acceptance and trust, which were also explicitly discussed by experts.

User Confusion consists of two aspects. First, the users might be confused about the vehicle behaviors that are not aligned with driver expectations. Second, users might get confused about the current mode being operated, which can lead to severe consequences. Both types of confusion could be derived from badly designed human-machine interfaces that lack explanations and salience when the automation systems take any actions or change their states.

Decreased User Acceptance and Trust. As a consequence of experiencing those badly designed interfaces and increasing reports on safety concerns regarding the automated vehicles, the general public's acceptance and trust in those advanced technologies has decreased, as stated by the experts, which further results in reduced technology adoption.

4.2.2.2 Theme 2: Explanation Mechanisms and Adaptability

Experts discussed what explanation is, how it works, and how it needs to be adapted accordingly. In general, an explanation is "a process" that takes human-AI interaction rather than a "state." While real-time explanation can facilitate understanding when interacting with automation systems, such a process does not necessarily intrigue explanation opportunities. Such opportunities occur when surprising events happen or when there is a change in the system's status. In those cases, query-based explainable systems can be used to understand users' explanation needs (e.g., "what drivers really want to see"). Generally speaking, explanations largely depend on context and situations. How explanations are presented depends on and is limited by the data used to train the explainable systems.

In line with the statement that "explanation is a process," experts also consistently agreed that **human-AI understanding is fostered through retrospective learning.** Users learned from their repeated experiences with "what it does well, what it does poorly, and circumstances in which it might fail completely." In other words, users understand the strengths and weaknesses of system capabilities

throughout this process, and they gradually understand system boundaries "after establishing basic reliability." In addition to their own experiences, users can also learn from tutorials and other users' experiences, such as the cautionary tale website offered by Ford. Throughout this learning process, users gain "predictability" and "tolerance" (for system inconvenience and annoyance) towards system behaviors. Such interaction processes can also influence trust, which is also a "process" rather than a state in which people want to "reach some calibration target," and there is no "ideal state of trust." Throughout repeated experiences, users might develop justified trust and justified mistrust while holding some unjustified trust and mistrust.

Experts discussed some examples of good explanations that promote positive user experiences, such as blind spot alerts, and rearview cameras. These systems did not require complicated explanations, but users were able to understand and make use of them. The experts claimed that such good experiences with well-explained systems could resolve uncertainty and confusion.

Four experts also mentioned that great individual differences exist when it comes to explanation needs and effectiveness. Such individual differences include but are not limited to expertise (expert vs. novice users), technology literacy, cognitive capabilities, curiosity, general attitudes toward the explanation, and cultural differences. Users who have more experience with automation systems after interacting with them for an extended period of time might no longer need explanations, while novice users might require extensive explanations to facilitate the learning process. People with better technology literacy might be more comfortable with interpreting sensor data and researching the system functionalities, while people who are not tech-savvy might develop different mental models. Users' cognitive capabilities might also influence their ability to comprehend and understand systems; "some are better, and some are not." In addition, some users are more "curious" or "skeptical" and thus seek additional information from the interface. In summary, effective explanations need to adapt to the person.

4.2.2.3 Theme 3: Methodologies, models, and frameworks for in-vehicle explanation design

All experts discussed existing methodologies, models, or frameworks that can be used to guide the in-vehicle explanation design. Most of the methodologies and models that the experts have discussed can be found in the existing knowledge body of human factors and cognitive engineering. For instance, the law of cognitive systems involving cognitive psychologists, cognitive task analyses, and the Situation Awareness-based Agent Transparency model (Chen et al., 2017) mentioned throughout the interviews were all established methods or frameworks that have been implemented extensively in different domains. Experts also claimed that basic human perception principles could also be referred to when designing explanations in the AV setting. The general user-centered design principles and processes would still be applicable to the in-vehicle explanation design. Relevant methods include but are not limited to user analysis, customer needs analysis, user testing field data analysis, heuristic evaluation, and adopting mixed-methods research approaches.

While the existing methodologies within the larger human factors and cognitive engineering domain could point out directions, experts also discussed the importance of borrowing techniques or methodologies from other related domains, such as computer science (XAI algorithms such as LIME, SHARP) (Islam et al., 2021) and aviation (Airforce technology), home entertainment, etc.

In general, the experts stated that the existing, well-established methods were all good resources to start with in-vehicle explanation design. While they were open to building specific frameworks for providing explanations, such work could be challenging.

4.2.2.4 Theme 4: Traps and tricks in in-vehicle explanations design

In addition to the methodologies and frameworks discussed to support in-vehicle explanation design, experts also mentioned traps and tricks when designing explanations that should be avoided or carefully considered.

Two experts mentioned that providing explanations does not necessarily need a display or interface. This might be a trap that "computer scientists" and researchers might fall into when they try to use new technologies to address issues with previous technology. The expert stated that "you don't necessarily want to have a display' and "the solution is not always just let's build more technology."

Two experts pointed out the issue regarding the current research paradigm. They argued that when designing and evaluating the explanation systems, experimental and cognitive psychologists were not involved completely or not early enough. A single experiment session can test only limited explanation options, which might not be sufficient to uncover the big picture. In the meantime, often in many studies, "the researchers never bothered to ask the participants questions about what they were thinking." Thus, user needs and their actual thought process were often overlooked. Additionally, the expert pointed out that a number of existing studies did not incorporate "large safety measurements" when evaluating explanations, which might miss the potential risks associated with providing explanations.

Three experts mentioned the traps associated with explanation mechanisms. They further articulated that not every piece of information needs to be explained. Users "don't need and want to have everything explained to me," and they may not "need to know every single detail," which can be distracting or annoying. Rather, the "profound details of how, when, and what will it work for" should be considered for certain situations. On the other hand, the experts mentioned that developing a comprehensive list of situations to be explained is not feasible because "situations are just too diverse." Especially with AI systems that rely on large-scale data training, "explainable AI may not be able to provide comprehensive explanations for complex decisions." Even with explanations, researchers should not assume "alert (vigilant) operators who are paying attention and have good situation awareness, which is unlikely to always be the case."

Last but not least, three experts also pointed out the conflicts between the needs of vehicle manufacturers and the needs of the end users, safety researchers, or government agencies. One of the experts pointed out

that “the marketing material presented to potential customers is designed to entice a purchase” rather than “reinforcing (and often contracting) the intended use or limits of the system.” Aligning with this, manufacturers also need to beat their competitors, which often results in user confusion and inaccurate mental models because of different “system naming.” Such inherent competition might yield isolated solutions and prevent generalizable ones that benefit end users.

4.2.2.5 Theme 5: Design considerations for in-vehicle explanation design

The experts discussed intensively design considerations for in-vehicle explanation design. The first step to designing a good explanation system is to clarify the goal or objectives of such systems. For any given system, the general function objective is to support user tasks. In the context of AVs, the objective is to support driver goals, which is safely traveling from point A to point B, with or without time pressure. Subsequently, the information provided by the systems is expected to facilitate “desired driver behaviors” and “not interfere with driver tasks and not overwhelm or distract drivers.” The experts emphasized facilitating shared “human-AI experience” where the machine should “augment” humans to make “informed decisions” rather than vice versa. Further, the experts described the goal of explanation as “bridging the gap between a complex system and human understanding.” To achieve this goal, the system could provide its “general awareness of the world” to indicate its context awareness. The goal of such systems is also to promote trust but “reflect risk accurately,” which makes human operators robust to system failures. In addition, different from most of the other AI systems, users for AVs are “inside the technology” and cannot “escape” easily, which makes “transparent communication” more critical due to safety concerns.

With a clear function objective, several design considerations were summarized as subthemes, ordered by frequency: (1) explanation design should consider driver-automation function allocation, (2) impacts of technological aspects on explanation design, (3) automation systems should show their understanding of human, and (4) temporal features of explanations.

Explanation design should consider driver-automation function allocation. Four experts mentioned that driver-automation function allocation needs to be considered when constructing explanations. The "division of labor," "expectations of the teammate as AVs," and "expectations to drivers" should be clearly communicated to avoid "driver confusion" during the "authority transition" stage. Even with a clear division of labor, the systems are still expected to support situation awareness and option awareness.

Impacts of technological aspects on explanation design. The experts also discussed the constraints that are put in explanations for technological limitations. Most AI systems, including AVs, operate on "a permanent sensor stream of data." In addition, different sensors might yield different decisions. These technological constructs make it challenging for users or even engineers to trace back a decision made by the vehicle. One of the experts also pointed out the limitations of these technologies in processing social information, which humans have processing advantages of. For instance, "humans have an innate ability to infer the intention of other drivers or pedestrians in a manner that physical sensors simply are not designed to do." All of these technological barriers can impact the actual implementations and effectiveness of explanations. The development of natural language processing (e.g., ChatGPT) that allows users to "probe a decision-making process" might provide a solution to address these challenges.

Automation systems should show their understanding of humans. The experts argued that to provide adaptable explanations, automation systems should also show some understanding towards humans. For example, understanding human capabilities and limitations, detecting human states, providing intervention, and supporting driver attention. While users are actively learning systems, the systems also need to learn user behaviors and patterns. The experts also mentioned the importance of "engaging users in the tasks regularly to ensure situation awareness" (e.g., providing options and asking users to confirm if not under time pressure) rather than waiting "until an error state." Querying whether users understand a certain situation is also helpful in establishing a system understanding of humans.

Temporal features of explanations. There are two main aspects of temporal features of explanations. First, time criticality was mentioned by two experts as the characteristic that differentiates explanations in the AV setting from other contexts. Thus, when presenting explanations, the decision-making mechanism needs to be considered under time pressure because everything happens "seconds to seconds" in the vehicle. Second, the explanation timing—ad hoc explanation vs. post hoc explanation—should also be taken into consideration along with the context.

Although the experts have discussed different aspects of design considerations, they also emphasized that there are tradeoffs inherent in explanation design, similar to selecting modalities to deliver explanations. There are no "one-fits-all" explanations that can suit all occasions.

With all of these discussions, one of the experts also posed that the explanations might not work due to the associated risks with providing explanations.

4.2.2.6 Theme 6: Risks associated with adding explanations

Seven experts discussed the risks associated with providing explanations. One of the most frequently mentioned risks was overtrust and overreliance. With the addition of explanations, users might develop unrealistic trust towards the systems and rely heavily on the systems to make decisions, which results in "cognitive laziness." Considering that people have a tendency to "rely on heuristics," "rely on our patterns," and "rely on our habits," it would not be a good situation "if we build up the expectation that the machines have all the answers." Drivers might be more likely to engage in secondary tasks such as looking at their phones, and they would have decreased situation awareness. Subsequently, such overreliance and overtrust can lead to skill degradation.

Experts also discussed extensively regarding the legal and ethical issues associated with providing explanations. They were concerned that providing the explanations would become the "get out of jail free card" as the automobile manufacturers would claim that explainability would predict a higher situation awareness. The liability determination can be challenging at that point. Clear regulations and guidelines

regarding how these explainable systems could and should be used might be a solution to mitigate this risk.

Other risks include that users might misunderstand explanations, which results in suboptimal decision-making, especially under time pressure. Misunderstanding could potentially further increase driver annoyance as the drivers get familiar with the driving automation systems, especially if the systems generate a lot of false alarms, which can be related to the current consequences of badly designed interfaces discussed in Section 0. As a consequence, users might abandon the systems and deactivate them.

4.2.2.7 Theme 7: Educating and training AV users

Three experts mentioned that educating and training, in addition to intuitive display design, could help mitigate the risks associated with introducing explanations. The form factors of such education and training could come from a variety of venues to ensure effectiveness. From the dealership aspect, they could provide interactive tutorials that immerse users in potentially problematic situations, which could help users develop a sense of "predictability" towards the technology. Such tutorials could also be made available online through user-initiated forums. Driving schools could also provide a certification process that ensures drivers are equipped with appropriate knowledge prior to being licensed to drive an AV. In line with this, car manufacturers could also establish a gamification program where certain features could only be unlocked after the users complete some part of the training. It is also recommended that the ongoing educational materials be available to users, and the educational and training materials should be updated as the technology advances.

4.2.3 Discussion

Findings from the expert interview delineated a larger picture of the current issues, challenges, and opportunities to develop explainable interfaces for AVs.

The consequences of a badly designed human-machine interface matched with what has been identified in contextual inquiries. The experts mentioned their own negative experiences that caused confusion, considering that they were also end users. Additionally, the experts were able to predict user attitudes based on their interaction with the technologies.

The major contribution of this series of expert interviews is to define explainability from multiple aspects and the adaptability requirements for the explanations. The experts emphasized heavily on the statement that "explanation is a process" and required continuous human-AI interaction through retrospective learning. This finding also matched with the results from contextual inquiries but might be often overlooked in current research studies where researchers expect a single experimental session to be able to determine the explainability of a certain design. In relation to the fact that explanation and understanding take time, training and educational materials should also support this learning process.

While individual differences were also identified through contextual inquiries, findings from the experts uncover specific areas that researchers can investigate while considering the adaptivity of explanations.

While some experts mentioned the possibility of having a new design framework for explanations in AVs, most experts expressed that existing design guidelines, models, and frameworks are still applicable to designing explanations for AVs after identifying the uniqueness of the current use context. Four components of human factors –human, task, tools, and environment—need to be considered altogether. From the human perspective, explanation designs require the system to show some level of understanding toward humans, including the individual differences listed in Section 0. The system also needs to consider the driver's state (e.g., attentive state and emotional state). From the task perspective, the design of an explanation needs to consider human-automation function allocation, which can interact with the driver state. From the tools' perspective, the design of the explanation needs to consider technological impacts, including technological availability and limitations. Finally, the environmental factor considers the

human-machine system and how AVs are different from other AI systems. Both time criticality and users-inside-technology specialize AVs from other AI-powered decision-support systems.

Although general user-centered design guidelines are still insightful in developing explanations for AVs, experts also raised the concern about falling into traps. Although several categories of traps were mentioned, the root cause of these traps is that user needs were overlooked or not considered properly, which has been known as automation abuse (Parasuraman & Riley, 1997). Whether it is not querying users about their thought processes during experiments or presenting unnecessary information that users might not need, either of these traps reflects the negligence of user inputs. In addition to the traps that researchers might fall into, these traps also reflect the conflict of interests that exist in vehicle manufacturers whose primary interest is to advertise their products rather than communicating system limitations that might potentially influence users' purchase decisions.

Finally, although providing proper explanations has a variety of benefits, the experts also discussed the risks associated with this addition from both users' and car manufacturers' perspectives. From the users' perspective, overtrust has always been an issue associated with advanced technologies, which have been discussed intensively in the past and recent publications (Kundinger et al., 2019; J. D. Lee & See, 2004; Xu, 2021). Users might have heuristic biases when they get accustomed to the explanations provided to them and rely on the patterns. Along with overtrust, badly designed explanations can cause misunderstanding and, subsequently, user confusion. From the vehicle manufacturer's perspective, no matter whether users receive good explanations and develop overtrust or receive bad explanations that yield confusion, the vehicle manufacturers might consider providing explanations as an "escape" from liabilities related to incidents, which should be taken great consideration and strict regulations should be developed to clearly communicate the responsibilities (Hubbard, 2018).

4.3 General Discussion

The findings from expert interviews support those identified in contextual inquires but in a more systematic and structured manner. Results from contextual inquiries and expert interviews both pointed out that the general display design principles are applicable in the context of designing explanations for AVs. Based on both qualitative studies, several in-vehicle explanation design guidelines are summarized below.

- **Explanations do not necessarily embed within a digital interface.**
    - Using natural language such as speech can be more efficient.
- **Consider human-automation function allocation and clarify responsibilities of both human operators and automated systems.**
- **Apply general display design principles (J. D. Lee et al., 2017b), especially mental model principles to design explanations.**
    - Pictorial realism: visual displays indicating continuous vehicle perception can increase drivers' confidence in AV performance. Proper use of pictorial realism can improve the efficiency and effectiveness of system explanations.
- **Follow user-centered design principles.**
    - Involve experimental psychologists, cognitive psychologists, or human factors experts.
    - Query user thought process rather than solely relying on subjective questionnaires to collect user evaluations.
    - Design information presented that support user goals while preventing information overload.
- **Involve experimental psychologists and cognitive psychologists early in the design cycle.**
- **Facilitate users retrospective learning through training, education, and repeated experiences of system boundary conditions.**

- **Establish clear regulations to mitigate the risks associated with unclear liability and ambiguous human-automation responsibility sharing.**

For the objective of this dissertation, I focus on designing explanations based on human-automation function allocation and adopting speech communication as a natural user interface.

CHAPTER 5 STUDY 5: DESIGNING AND EVALUATING EXPLAINABLE IN-VEHICLE INTELLIGENT AGENTS

5.1 Explainable In-Vehicle Intelligent Agent Design

Based on the existing literature and findings gathered from Study 3 and Study 4, three explainable in-vehicle intelligent agent designs were proposed. Specifically, I incorporated Lyons' Models of Transparency for Human-Robot Interaction (HRI) (Lyons, 2013) while proposing agents to promote drivers' situation awareness (SA) towards the automation system and the environment. It is noted that the term "robot" in Lyons' model refers to any system or agent that operates under a certain degree of autonomy (Lyons, 2013).

5.1.1 Lyons' Models of Transparency

Lyons claimed that the transparency situated in HRI consists of two pieces of information that (1) a robot system needs to convey to a human and (2) the system needs to express awareness or understanding of human states (Lyons, 2013). The former one is robot-to-human transparency, and the latter is robot-of-human transparency. Both factors need to be considered for a well-designed explainable interface.

The robot-to-human transparency involves four models: the intentional model, the task model, the analytical model, and the environment model. Based on the intentional model, the system needs to communicate its design, purpose, and intent (Lyons, 2013)—which refers to a higher level of understanding of the system's purpose and how it serves its purpose. The task model suggests communicating the system's understanding of the current task, its intent in regard to task-related goals to be accomplished, its progress in achieving these goals, and the system's awareness of its capabilities in the current context (Bhaskara et al., 2020; Lyons, 2013). The analytical model advocates communication regarding the underlying analytical principles used by the system to make decisions. In other words, the analytical model deals with the system's decision-making process, for instance, what information acquired from which sensor helps with coming up with a decision through what algorithm. Most of the

explainable machine learning methods are closely related to the analytical model to make the "black boxes" more transparent and interpretable. Finally, the environment model urges the system to communicate its capability of understanding the dynamics of the surrounding environment where it is operating, such as geographic variance, weather conditions, and potential threats in the proximity (Lyons, 2013).

The robot-of-human transparency includes the teamwork model and the human state model. The teamwork model emphasizes a clear understanding of human-automation function allocation from both parties, the system and the human. The system should convey clearly the responsibilities held by itself and the human operator, along with its current level of autonomy (Lyons, 2013). Once the shared awareness of responsibility allocation is established, the human state model raises the importance of the system communicating its understanding of the humans' cognitive, emotional, and physical state (Lyons, 2013). The robot-of-human transparency points out the significance of adaptive automated systems.

5.1.2 Principles of Designing for Situation Awareness

In addition to the model of transparency, the system should also consider drivers' situation awareness (SA) to complete the driving task. Study 2 presented in Chapter 3 indicates that drivers might already have good lower level SA. In addition, presenting lower level data requires human operators to calculate or evaluate information (Endsley, 2016). This additional mental operation slows down the response selection stages of human information processing, which can be critical in the driving context where imminent threats are in proximity. Thus, directly supporting higher level SA in the driving context might be able to support drivers' decision-making process without overloading them with unnecessary information. Endsley (2016) has proposed eight principles of designing for SA. For the scope of this dissertation, I mainly focused on the first three principles that are highly relevant to the topics of interest.

Principle 1: Organize information around goals

Endsley (2016) emphasized that the information provided to human operators should be centered on their major goals. Under the driving context, drivers' goal is to travel from position A to position B safely and timely. Thus, in my study, the information focuses on drivers' primary driving task as safe driving and does not involve information related to their secondary driving tasks (e.g., using indicators) or tertiary driving tasks (e.g., manipulating air conditioners). Following this principle can also prevent information overload.

Principle 2: Present Level 2 information directly – Support comprehension

This principle suggests that directly presenting information to support comprehension can be efficient in some cases to remove the extra process of integrating cues and formulating meaningful information, which can be beneficial to novice users. Endsley (2016) also indicates that sometimes only Level 2 SA is necessary and lower Level 1 data might not be required—which represents the current user context that drivers might already have good lower Level 1 data as indicated in the findings in Chapter 3, Study 2.

Principle 3: Provide assistance for Level 3 SA projections

The third principle advocates for assisting Level 3 SA projections, which is one of the most challenging and effortful parts of SA, especially for novice users who typically lack a well-developed mental model (Endsley, 2016). In the driving context, providing predictive aid (J. D. Lee et al., 2017c) can support Level 3 SA and assist drivers to plan their maneuvers.

5.1.3 Explainable Intelligent Agent Design

With these two frameworks and the findings from Chapter 3 and Chapter 4, I proposed alternatives of explainable intelligent agent that can provide environmental and system information to support drivers' primary driving tasks in conditionally automated driving. The information structure of the proposed agents was developed based on Lyons' Model of Transparency and three principles of designing for SA.

Incorporating three principles of designing for SA, the intelligent agents proposed in this dissertation mainly focused on supporting drivers' Level 2 and Level 3 SA. Subsequently, according to the teamwork model and the design guideline that explanations should consider human-automation function allocation, I proposed that the information provided by the agent varied based on whether the driver was in control of the vehicle. When the driving automation system is in control, drivers are not actively engaging in the driving task and thus might have degraded SA (De Winter et al., 2014). Therefore, both Level 2 and Level 3 information is provided to them to keep drivers in the loop. While drivers are in control and actively seeking information to respond to the road events, they in general have better SA, assuming model drivers (e.g., attentive, safe drivers). In this case, I hypothesize that drivers under manual driving will require less information and system support, which motivates the adaptivity of explanations: providing only either Level 2 or Level 3 SA information to drivers in control.

Table 5.1-1presents an example of the intelligent agent design according to the models and principles discussed above. A full list of messages used in this dissertation can be found in Appendix D.

Table 5.1-1 Example agent explanation design.

Automation Mode	Agent 1	Agent 2	Agent 3
Autonomous	*Slow moving car ahead, I will change lanes and pass it in a few seconds.*		
Manual	[Level 2 + Level 3 SA] *Static obstacles in your lane ahead, your lane will be blocked in 600 feet.*	[Level 2 SA] *Static obstacles in your lane ahead.*	[Level 3 SA] *Your lane will be blocked in 600 feet.*

Based on the existing evidence, I further have the following hypotheses:

- **H1**: Participants' subjective evaluations will favor the agent providing more information.
- **H2**: Participants' driving performance will indicate that agents providing less information yield better performance.

5.2 Method

5.2.1 Participant

For a within-subjects design with one independent variable of three levels, to observe a medium effect size for the variable of adaptive agent type with a power of 0.8 and an alpha error probability of 0.05, a total of 28 participants are needed.

Participant recruitment information was sent out after the study protocol was approved by the Virginia Tech's Institutional Review Board (IRB# 24-245). A total of 41 participants were recruited, from which two participants were excluded from the data analysis due to not following the instructions. The remaining 39 participants (11 females) aged between 19 to 30 years old (*Mean* = 22.15, *SD* = 2.82) with an average driving experience of 5.38 years (*SD* = 2.90).

5.2.2 Apparatus and Stimuli

A motion-based driving simulator (Nervtech^{TM}, Ljubljana, Slovenia) was used in this study. The simulation driving scenarios in this study were developed using the SCANeR studio to include highway (speed limit: 110 - 130 km/h, or 68.4 - 80.8 mph), rural (speed limit: 70 - 90 km/h, or 43.5 – 55.9 mph), and city (speed limit: 50 km/h, or 31.1 mph) areas. The experimental vehicle was designed to simulate an SAE Level 3 Conditional Driving Automation (SAE International, 2021).

5.2.3 Experimental Design

A within-subjects design with the adaptive intelligent agent type (non-adaptive, support Level 2 SA when manual, or support Level 3 SA when manual) was adopted. Participants experienced all types of adaptive intelligent agents in three drives. The order of interfaces presented and the matching between the conditions the scenarios were counterbalanced using Latin-square. Thus, the same agent was not always used in the same scenario to avoid the scenario as a confounding variable.

Each driving scenario started with autonomous driving and ended with manual driving. Drivers experienced autonomous driving and manual driving segments interchangeably, resulting in four segments in total (Figure 5.2-1). The intelligent agent first introduced herself and clarified her

responsibilities as well as the driver's ones. In each driving scenario, there were a total of eight road events (see Appendix D). Half of the events occurred while the driving automation system was engaged. The intelligent agent provided explanations eight seconds prior to taking actions. Both pieces of information related to Level 2-Comprehension and L3-Projection were provided. The drivers also went through two non-urgent takeover situations, with a lead time of 12 seconds, which fell under the range of most appropriate takeover lead time based on subjective ratings (i.e., 12-18s) for non-urgent events such as manual highway exit (Tan & Zhang, 2022). During manual driving segments, the agent provided information eight seconds prior to the onset of the event to support their Level 2 SA, Level 3 SA, or both, corresponding to road events. Eight-second was selected because it yielded the lowest crash rate when drivers were alerted about the road hazard by speech warnings (Y. Zhang et al., 2016). After the first manual driving segment, there was a handover request where drivers were asked to transition the vehicle control back to the automation system.

5.2.4 Procedure

Upon arrival at the lab space, participants were briefed about the experimental procedures and asked for written consent. Then, participants had a 5-minute practice session as a simulation sickness test run and for them to get familiar with the simulator scenario, simulator control, and study procedures. Before and after the practice, participants filled out simulation sickness questionnaire. Participants without simulation sickness tendency continued the study with a pre-experiment questionnaire that collected their demographics and their propensity to trust (Merritt et al., 2013).

Then, participants were guided to the driving simulator to drive through three scenarios in sequence, each lasted for about 10-15 minutes. Figure 5.2-1 shows an example of the driving scenario structure. During each drive, participants encountered eight events in total. After each event, the intelligent agent provided speech information to explain the situation. The events and agent explanations can be found in Appendix D. Then, participants answered several questions (see Table 5.2-2) about their perception towards the explanation. After completing each drive, participants completed a survey (see Table 5.2-2) capturing

their perceptions towards the intelligent agent they just experienced. Participants also verbally expressed their opinions towards the intelligent agent through a structured interview (see Appendix E). After participants completed all drives, they ranked the agent preference and explained their reasons behind the ranking.

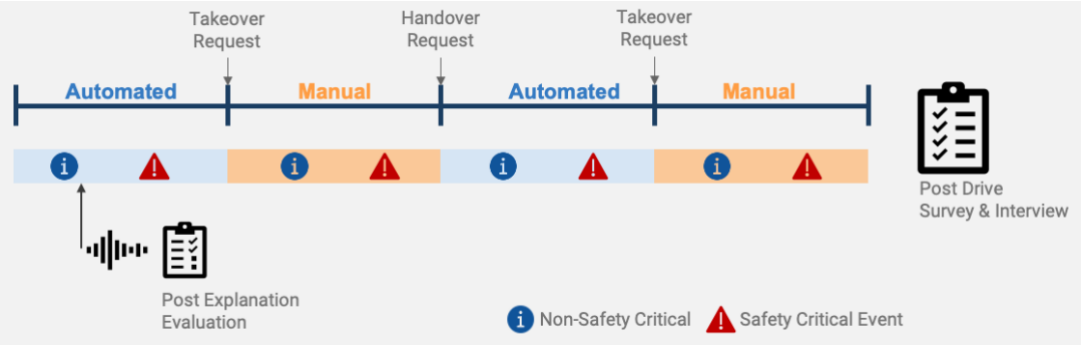

Figure 5.2-1 Example driving scenario structure.

5.2.5 Dependent Measures and Data Analysis

Objective and subjective measures were collected to evaluate drivers' perception towards the intelligent agents and their behavioral responses to agents' intervention. Objective measures consisted of driving behaviors from different aspects, including driver's takeover time, driver response time, and their driving performance for each event during the manual driving segments depicted in Figure 5.2-1. The takeover time was the interval between the agent issued the information for authorization transition and the driver take over the control. The driver response time is the time interval from the onset of the speech explanation provided by the agent to the earliest indication that a driver movement response initializes (SAE International, 2023). As for time-to-steer, considering that not all road events required steering response, only those required steering responses were included in the calculation. The driving performance included the measures for drivers' longitudinal control (e.g., speed) and lateral control (lateral acceleration and standard deviation of lane position) (McDonald et al., 2019). Specific metrics and their definitions can be found in Table 5.2-1.

Table 5.2-1 Driving performance measures and definitions.

Category	Measures	Unit	Definition
Driver Response Time	Time to Brake	seconds	Brake reaction time measured from the onset of the agent speech to when the driver pressed the brake pedal (SAE International, 2023).
	Time to Steer	seconds	Steering reaction time measured from the onset of the agent speech to when the steering wheel angle has changed more than 2˚, either to the left or right (SAE International, 2023).
	Time off Accelerator	seconds	Accelerator reaction time measured from the onset of the agent speech to when the driver completely released the accelerator pedal (SAE International, 2023).
Driving Performance	Speed	m/s	Maximum/Minimum/Average speed and standard deviation of the speed during the manual driving time period of a TOR event.
	Lateral acceleration	m/s^2	Maximum/Minimum/Average lateral acceleration during the manual driving time period of a TOR event.
	Standard deviation of lane position (SDLP)	meters	The standard deviation of the lateral distance of the subject vehicle regarding the middle of the lane traveling, calculated across the manual driving time period of a TOR event.
	Jerk	m/s^3	Maximum/Minimum/Average jerk, which is the time derivative of vehicle acceleration. Positive values indicate vehicle acceleration and negative values indicate vehicle deceleration. .

Table 5.2-2 presents a summary of subjective dependent measures using validated questionnaires collected throughout the procedure described in Section 5.2.4. After participants completed all drives, their preference ranking toward the agents was collected, along with their rationale for the ranking.

Table 5.2-2 List of dependent measures and their definition.

Time of Measurement	Measure	Description	Items and Instrument Type
Pre-Experiment	Propensity to Trust (Merritt et al., 2013)	"a stable, trait-like tendency to trust or not trust others" (Merritt & Ilgen, 2008, p. 195)	5-point Likert scale on 6 items
Post-Event	Situational Trust Scale (Holthausen et al., 2020)	The impact of contextual differences on trust development	7-point Likert scale on the Trust and Performance item
	Explanation Satisfaction scale (Hoffman et al., 2023)	The degree to which users feel that they sufficiently understand the automation system.	5-point Likert scale on 3 selected items.

Time of Measurement	Measure	Description	Items and Instrument Type
Post-Drive	Driver Activity Load Index (DALI) (Pauzié, 2008)	An instrument that is used to evaluate drivers' mental workload while completing driving activities, with or without the support from in-vehicle systems (e.g., intelligent agent)	Six dimensions rated on the 100 scale.
	Trust Scale for the XAI Context (Hoffman et al., 2023)	An attitude that the automation system will help achieve the driver's goals in a situation characterized by uncertainty and vulnerability (J. D. Lee & See, 2004).	5-point Likert scale on 8 items.
	Explanation Satisfaction scale (Hoffman et al., 2023)	The degree to which users feel that they sufficiently understand the automation system.	5-point Likert scale on the full 7 items.
	GodSpeed (Bartneck et al., 2009)	Including five factors: **Anthropomorphism**: attribution of human in nonhuman agents. **Animacy**: agent lifelikeness. **Likeability**: positive first impressions. **Perceived Intelligence**: perceived ability to generate intelligent and human-like behaviors. **Perceived Safety**: perceived level of danger when interacting with the agents.	5-point semantic differential scale on 24 items
	Robotic Social Attributes Scale (RoSAS) (Carpinella et al., 2017)	Including three factors: **Competence**: the intelligence or ability of the agent. **Warmth**: including factors of feeling, happy, organic, compassionate, social, and emotional. **Discomfort**: agent awkwardness.	7-point Likert scale on 18 items.

For continuous measures, a one-way repeated measures analysis of variance (ANOVA) was adopted to understand if the adaptive intelligent type had influences on dependent variables. If the assumption of sphericity was violated based on Mauchly's test of sphericity, Greenhouse-Geisser correction was adopted when reporting the results. If significant main effect was identified, pairwise comparison with the Bonferroni correction (*$p < .017$, **$p < .003$, ***$p < .0003$) was conducted to further understand the relationship across the levels. In addition, for post-event questionnaires, a paired-samples *t* test was used to understand whether the driving automation mode has an effect on driver's perception toward the same information structure (Level 2 and Level 3 information combined). A Chi-square test for each preference rank was conducted to identify differences in preferred agents.

5.3 Results

5.3.1 Driving Behavior

No collision was observed during the study. Table 5.3-1 presents the descriptive statistics corresponding to each category of driving behavior measures. The statistical analysis was reported in the subsequent sections. It is noted that drivers might adopt different response strategies towards road events; thus, the number of participants included in the analysis for each behavioral measure might vary.

Table 5.3-1 Descriptive statistics of driving behaviors [Mean (SD)].

		Intelligent Agent Type (i.e., information provided under manual driving)		
Category	**Measures**	**Level 2 + Level 3**	**Level 2**	**Level3**
Takeover Performance	Takeover Time [s]	7.39 (0.96)	7.35 (0.87)	7.44 (1.19)
Driver Response Time	Time to Brake [s]	5.18 (1.70)	4.93 (2.09)	4.39 (1.58)
	Time to Steer [s]*	6.15 (2.98)	4.89 (2.04)	4.67 (2.21)
	Time off Accelerator [s]	1.80 (1.01)	1.68 (1.21)	2.06 (1.47)
Driving Performance	maxSpeed [m/s]	24.99 (1.70)	25.19 (2.11)	25.24 (2.30)
	minSpeed [m/s]	13.34 (2.25)	13.98 (2.19)	13.85 (2.63)
	avgSpeed [m/s]	19.47 (1.65)	19.66 (2.03)	19.88 (2.48)
	max Lateral Acceleration [m/s^2]	1.71 (0.64)	1.78 (0.81)	1.95 (0.87)
	min Lateral Acceleration [m/s^2]	-1.93 (0.65)	-1.92 (0.46)	-2.03 (0.98)
	avg Lateral Acceleration [m/s^2]	-0.09 (0.28)	-0.10 (0.27)	-0.02 (0.32)
	SDLP [meters]	0.36 (0.08)	0.37 (0.09)	0.39 (0.11)
	maxJerk [m/s^3]	85.41 (17.10)	88.04 (15.98)	114.20 (85.97)
	minJerk [m/s^3]	-64.16 (16.37)	-63.97 (16.65)	-87.21 (88.45)
	avgJerk [m/s^3]*	-0.00 (0.02)	0.00 (0.03)	-0.02 (0.04)

* $p < .017$

5.3.1.1 Takeover time

Results from repeated measures ANOVA indicated that there was no significant main effect of intelligent agent type on takeover time, F (2, 74) = 0.22, p = .268.

5.3.1.2 *Driver response time*

There was a significant main effect of intelligent agent type on driver's time to steer, F (2, 74) = 4.29, p = 0.006, η_p^2 = .104. Pairwise comparison with Bonferroni correction indicated that drivers took longer time to steer when the agent provided both Level 2 and Level 3 SA information, compared to when the agent only provided Level 3 SA information (p = .0009) (see Table 5.3-1). There were no differences in the time to steer between the agent only providing Level 2 SA information and the agent providing both Level 2 and Level 3 SA information (p = .024), or the agent with Level 3 SA information (p = 0.33).

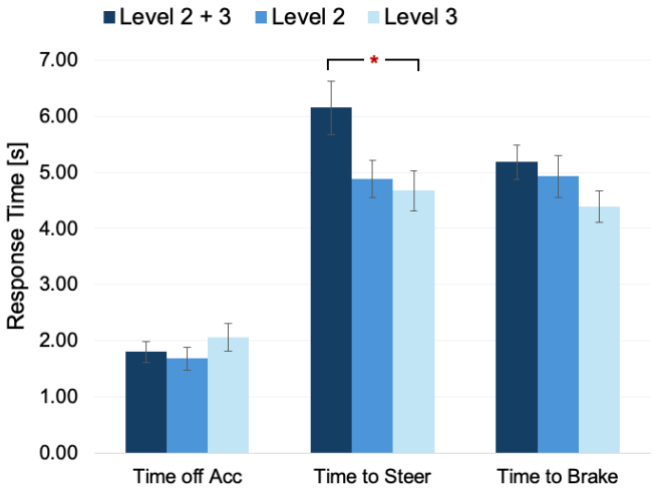

Figure 5.3-1 Driver response time across three conditions (Bonferroni Correction: *p < .0.017)
[error bars indicate standard errors]

No significant main effect of intelligent agent type was found on driver's time to brake: F (2, 60) = 1.53, p = .075, or time off accelerator: F (2, 68) = 0.98, p = .127.

5.3.1.3 *Driving performance*

There was a significant main effect of agent type on driver's average jerk, F (2, 74) = 3.96, p = 0.008, η_p^2 = .097. Pairwise comparison with Bonferroni correction indicated that participants had larger average jerk when provided only Level 2 SA information, compared to only Level 3 SA information (p = .011) (see Table 5.3-1). There were no differences in average jerk between the agent providing Level 2 and Level 3 SA information and agent providing only Level 2 information (p = 0.333) or Level 3 information (p

= .0.068). No significant main effect was found on the maximum jerk: F (2, 74) = 3.66, p = 0.020, or minimum jerk: F (2, 74) = 2.53, p = 0.039.

The intelligent agent type did not have significant main effects on any speed-related measures [maximum speed: F (2, 74) = 0.33, p = 0.241, minimum speed: F (2, 74) = 1.58, p = 0.071, average speed: F (2, 74) = 0.97, p = 0.129], lateral acceleration-related measures [maximum lateral acceleration: F (2, 74) = 0.96, p = 0.129, minimum lateral acceleration: F (2, 74) = 0.343, p = 0.331, average lateral acceleration: F (2, 74) = .645, p = 0.142], or SDLP [F (2, 74) = 0.84, p = 0.144].

5.3.2 Subjective Measures: Driver-Agent Interaction

5.3.2.1 Post-event measures

Results from the paired-samples t test indicated that participants had higher situational trust towards agent explanations (with Level 2 and Level 3 SA information) during autonomous driving compared to manual driving, t(38) = 2.04, p = 0.49, Cohen's d = 0.326. With the same information structure (i.e., Level 2 and Level 3 information), participants also perceived that explanations were more helpful to their goals in autonomous driving mode compared to manual driving mode, t(38) = 2.26, p = 0.03, Cohen's d = 0.361 (Figure 5.3-2). Participants did not have a significant difference in their agreement on the statements of "the explanation is satisfying" (p = .538) and "the explanation provides sufficient details" (p = .566).

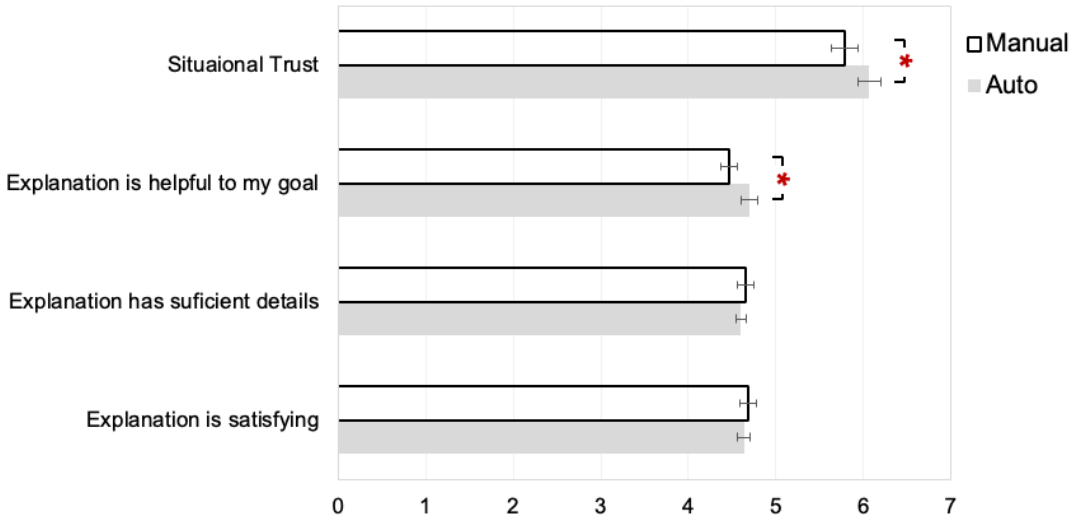

*Figure 5.3-2 Post events evaluations on agent explanations (level 2 + 3) between automated and manual driving. (*p < .05) [Error bars indicate standard errors]*

Figure 5.3-3 compares drivers' perception toward explanations after each event during manual driving. Through one-way repeated measures ANOVA, we found that the agent type has significant main effect on participants' situational trust: F (2, 76) = 5.86, p = 0.001, η_p^2 = .134. Pairwise comparison indicated that drivers have higher situational trust when the agent provided both Level 2 and Level 3 SA information compared to providing only Level 2 information (p = 0.006), or only Level 3 information (p = 0.008). The difference in the situational trust between providing Level 2 SA and providing Level 3 SA was not significant (p = 0.333). Results also indicated that there was a significant main effect of agent type on whether they thought the explanation was satisfying [F (2, 76) = 15.976, p < .0003, η_p^2 = .296], had sufficient details [F (2, 76) = 24.482, p < .0003, η_p^2 = .392], and was helpful to the goals [F (2, 76) = 10.703, p < .0003, η_p^2 = .220]. Regarding whether the explanation was satisfying and provided sufficient details, pairwise comparisons with the Bonferroni correction indicated that participants rated the highest towards the explanation with both Level 2 and Level 3 SA information, followed by only Level 2 SA information, and rated the lowest towards only providing Level 3 SA information (Figure 5.3-3). Regarding "helpful to the goals", participants rated the explanation with Level 2 and Level 3 SA information higher than only with Level 2 SA (p = .001) or Level 3 SA (p < .0003); however, the differences between the Level 2 and Level 3 was not significant (p = 0.043).

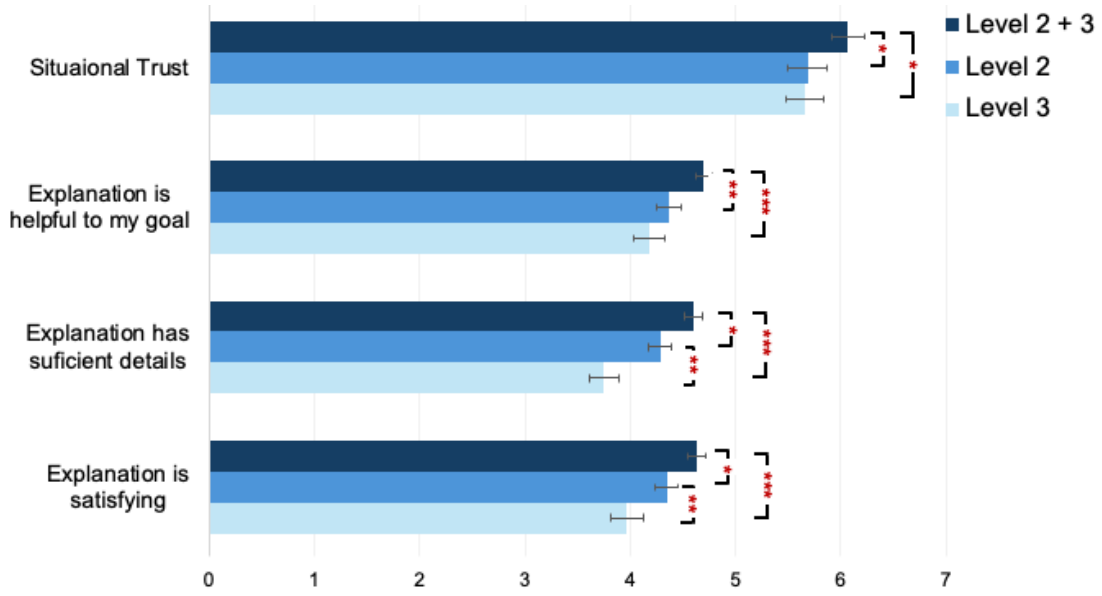

*Figure 5.3-3 Post event evaluations on agent explanations across three conditions. (Bonferroni Correction: *p < .0.017, **p < .003, ***p < .0003) [error bars indicate standard errors]*

5.3.2.2 Overall evaluation toward three types of intelligent agent

Table 5.3-2 shows the descriptive and inferential statistics of each questionnaire used to evaluate the overall experiences with three types of intelligent agents. Results from repeated measures ANOVA indicated that the agent type did not have a significant main effect on any of the dimensions in the subjective questionnaires.

Table 5.3-2 Descriptive and inferential statistics of subjective ratings on overall agent evaluations. [Mean (SD)]

		Leve 2 + 3	Level 2	Level 3	F statistics
DALI		58.40 (16.48)	54.40 (15.44)	53.08 (15.47)	F (2, 74) = 2.64, p = 0.026
Trust in XAI		3.81 (0.56)	3.86 (0.58)	3.83 (0.53)	F (2, 76) = 0.47, p = 0.309
Explanation Satisfaction		3.45 (1.09)	3.54 (1.06)	3.47 (1.08)	F (2, 76) = 0.74, p = 0.160
GodSpeed	Anthropomorphism	3.53 (0.81)	3.41 (0.85)	3.36 (0.94)	F (2, 76) = 1.82, p = 0.060
	Animacy	3.49 (0.82)	3.44 (0.90)	3.49 (0.94)	F (2, 76) = 0.19, p = 0.274
	Likability	3.86 (0.49)	3.88 (0.51)	3.93 (0.62)	F (2, 76) = 0.52, p = 0.195
	Perceived Intelligence	4.22 (0.57)	4.27 (0.61)	4.26 (0.60)	F (2, 76) = 0.27, p = 0.254
	Perceived Safety	3.69 (0.72)	3.79 (0.74)	3.85 (0.63)	F (2, 76) = 1.63, p = 0.067
RoSAS	Competence	5.74 (0.99)	5.68 (1.03)	5.57 (1.05)	F (2, 76) = 1.42, p = 0.083

	Leve 2 + 3	Level 2	Level 3	F statistics
Warmth	3.03 (1.49)	3.05 (1.38)	3.03 (1.35)	F (2, 76) = 0.02, p = 0.326
Discomfort	1.44 (0.46)	1.48 (0.71)	1.43 (0.60)	F (2, 76) = 0.10, p = 0.303

5.3.2.3 Agent preference ranking

Table 5.3-3 presents the preference ranking and distribution for each type of agent. Participants' preferences were significantly different across three types of agents, $\chi^2(4) = 32.57$, $p < .001$. Further analysis indicated a significant difference in participants' 1st preferred agent was found: $\chi^2(2) = 10.87$, $p = .004$. The non-adaptive agent (i.e., providing Level 2 and Level 3 SA information in both automation modes) was preferred the most.

Table 5.3-3 Preference ranking for all agent conditions.

Preference	Agent Type		
	Level 2 + Level 3	Level 2	Level 3
1st	**21**	14	4
2nd	15	15	9
3rd	3	11	25
Average Score	1.15	1.97	2.49

Note: unit – number of participants

When participants explained the reasons behind their first preferred agent, the perceived information value was the most frequently mentioned factor (N = 20), followed by perception (N = 9) and amount of information (N = 5). Information value refers to that explanations were perceived to be helpful to driver goals and could support drivers' situation awareness, especially objects that were not visible to drivers for different reasons (e.g., fog, blind spot, etc.). Perception refers to the subjective evaluations resulted from driver-agent interaction. The amount of information simply refers to how much information the agent provided. Table 5.3-4 provides some examples for each factor.

Table 5.3-4 Participants' reasons regarding their most preferred agents.

Category	Direct Quote Examples from Participants
Information Value	*"It told me things that I would have like not seen" (P2, prefer Level 2 + Level 3)* *"It informed me of what was in the environment as well as telling me what actions I could take or should take, especially moments when I was driving." (P27, prefer Level 2 + Level 3)* *"(It was) knowledgeable about the surroundings and how to handle them." (P16, prefer Level 2)*
Perception	*"(It) just felt like a bit more natural and kind, I want to say; it was just more audibly appealing." (P5, prefer Level 2 + Level 3)* *"I felt it was a bit comfortable." (P21, prefer Level 2 + Level 3)* *"The system was more competent." (P3, prefer Level 2)*
Amount of Information	*"I also felt like it provided me with the most amount of information." (P18, prefer Level 2 + Leve l3)* *"It's accurate and it gave a lot of information." (P28, prefer Level 2)*

There was a significant difference in participants' least preferred agents: $\chi^2(2) = 20.11$, $p < .001$. When participants were asked to reason their least preferred agents, information value ($N = 9$) was the most frequently mentioned factor, followed by perception ($N = 8$). Table 5.3-5 presents some examples within each category.

Table 5.3-5 Participants' reasons regarding their most preferred agents.

Category	Direct Quote Examples from Participants
Information Value	*"Didn't tell you really that many details about anything." (P31, least prefer Level 3)* *"I feel like it didn't provide me as much information as I needed to help me drive safely." (P23, least prefer Level 3)*
Perception	*"First one kind of gave me stress." (P19, least prefer Level 3)* *"It's just the most forgettable one." (P5, lease prefer Leve l3)*

5.4 Discussion

In the context of conditionally automated driving, this study explored three types of adaptive intelligent agents that provided Level 2 and Level 3 SA information when the vehicle was driving, and provided Level 2 and Leve 3 SA, only Level 2 SA, or only Level 3 SA when the driver was driving. Drivers' driving behaviors and their subjective evaluations in response to the agent information were collected. Results indicated that driving behavior and drivers' overall evaluations toward three types of agents did not differ significantly. However, drivers' steer response time and their subjective ratings on the explanations for each event differed across three conditions. In general, drivers still preferred to receive both Level 2 and Level 3 SA information and were dissatisfied with the agent who only provided Level 3 information.

5.4.1 Drivers still prefer agents providing comprehensive information

Results from situational trust and explanation satisfaction for each individual event indicate that providing both Level 2 and Level 3 SA was still favorable to participants, which supports my H1. In fact, previous studies also showed that drivers tend to prefer receiving more information. Zang (2023) adopted the situation awareness-based agent transparency model (Chen et al., 2017) and developed three types of in-

vehicle intelligence with increased transparency. She found that agents providing both Level 1 and Level 2 transparency yielded higher cognitive trust, lower workload, and better situation awareness compared to the agents providing only Level 1 transparency or all three levels of transparency (Zang, 2023), indicating that providing all information might have reverse effects on drivers in automated vehicles.

While drivers in this study favored the agents providing both Level 2 and Level 3 information, they also complained about receiving unnecessary information from that type of agent. Thus, finding the balance between providing valuable information and not overloading drivers requires further research, as discovered in the expert interviews.

5.4.2 Consistency between driving performance and subjective evaluations

Considering that the three types of agents only differ when they present information under manual driving, some participants were not able to indicate such subtle differences explicitly. Thus, most of the participants' driving performance metrics did not reach significant differences across different levels, and so did most of the subjective evaluations on the overall experiences with the intelligent agents. This unperceivable difference might have the benefit of reducing the sense of inconsistency.

Participants had a larger negative jerk during manual driving when accompanied by the agent only providing Level 3 information. In general, a larger jerk is highly correlated with greater driver stress (Othman et al., 2008). In addition, a larger negative jerk has been found to be highly related to aggressive driving behaviors (Feng et al., 2017). Thus, drivers accompanied by agents providing Level 3 information might be experiencing greater stress after handling road events and exhibiting some aggressive driving behaviors. This finding is consistent with the preference ranking, which showed that participants did not favor the agent who only provided Level 3 SA information. Participants reported that they got "surprised" and "confused" when hearing the message from the agent with Level 3 SA and then started to look for objects of interest on the road, which might explain the increased stress experienced by participants and subsequently lead to greater negative jerks.

5.4.3 Driver response time indicates an advantage for less information

No significant difference in drivers' brake response time was identified across three conditions. However, the steering response time indicated that drivers receiving both Level 2 and Level 3 information had significantly longer steering response time compared to them receiving only Level 3 information and showed a tendency to be longer than them receiving only Level 2 information.

It has been known that braking, rather than steering, is the most common maneuver for drivers in traffic conflicts, especially in crash-imminent events (Abraham et al., 2018; Dozza, 2013). Given that alerts were issued eight seconds prior to the points of interest on the road, events in this study were not necessarily considered crash-imminent events. Thus, drivers had a longer time to make brake-or-steer decisions under relatively lower time pressure, which could be referred from the longer brake response time (*Mean* = 4.83 s, *SD* = 1.76 s) and steering response time (*Mean* = 5.24 s, *SD* = 2.58 s) compared to those found in Dozza (2013) where near-crash events with greater safety criticality were analyzed. With longer decision-making time, drivers' brake response time was still slightly faster than the steering response time, which is aligned with the findings in previous studies that brake response time is, in general, faster than steering response time (Dozza, 2013).

The significant differences in the steering response time might indicate that drivers receiving less information were able to make steering decisions and respond faster if needed. It has been recommended that at least 3 seconds should be reserved for drivers to respond to the road changes (e.g., a door open) (Summala, 1981). With the eight-second lead time for the onset of obstacles, a faster steering response allows drivers a longer time to avoid potential collisions with safe operation. This result indicates the benefits of having less information in terms of steering response. In addition, considering the non-significant differences in brake response time, I would not contribute to such an advantage to the difference in message length. Results from driving performance and driver reaction time indicate that my H2 was partially supported. This finding might also indicate that providing explanations can alter drivers' automatic decision-making process to avoid heuristic biases. Based on the cognitive controlling of driving

(Michon, 1985) and Rasmussen's SRK (Skill, Rule, Knowledge) levels of cognitive control (Rasmussen, 1983), skill-based behaviors are involved in basic vehicle handling (operational level) and tends to develop automaticity; rule-based behaviors are at the tactical level that involves navigating through other vehicles and road users; and knowledge-based behaviors are more relevant to strategic driving control such as driving in unfamiliar situations (Hale et al., 1990; Ranney, 1994). Typically, hazard perception and response requires controlled, conscious processing (Ranney, 1994) that might involve knowledge-based behaviors. However, under time pressure, skill-based and rule-based behaviors are more likely to be activated and drivers tend to rely on experiences to make decisions, which is faster but does not always yield better or correct decisions due to heuristic biases such as availability of actions (J. D. Lee et al., 2017a). Adding explanations can potentially support drivers' knowledge-based behaviors by assisting different stages of decision making (J. D. Lee et al., 2017a). For instance, Level 2 SA information can support information interpretation and assessment as the second stage of decision making to reduce the information access cost, which can be greatly helpful for novice drivers. Level 3 SA information can supplement the plan and choose stage by providing alternative responding options to drivers. However, adding both information might prolong the information processing time, and thus slows down the decision-making process for knowledge-based behaviors. Therefore, the balance between adding explanations to support vs. prolong the decision-making process still needs further exploration.

The findings in driver response also indicate that there might be an incompatibility between subjective and behavioral reactions to different types of explanations. Although less information might speed up their response selection process, drivers subjectively did not feel comfortable with less information.

5.4.4 Contributions, Limitations, and Future Work

This simulator study integrated both Lyons' Model of Transparency (Lyons, 2013) and Endsley's Principles of Designing for Situation Awareness (Endsley, 2016) and designed three types of intelligent agents varied in explanations given to drivers during automated and manual driving based on the driver-automation function allocation. Results indicate that although drivers still prefer to receive more

information, their response selection in reaction to road events might receive more benefits with less information if designed properly.

This study has the following unique contributions. First, this is the first study that examined an adaptive explanation strategy based on driver-automation function allocation. Although overall subjective ratings did not show any significant differences, participants were still able to evaluate the explanations on a case-by-case basis (i.e., post-event evaluation). Second, the findings from this study indicate that explanations might influence drivers' response selection and reaction time, which serve as safe driving behavior measures that were overlooked as identified by experts. The modification of drivers' response selection resulted from providing explanations further suggest that proper explanations can support drivers' decision-making process and prevent heuristic biases. Finally, this study followed the guidelines of "querying users" mentioned in the expert interviews, which provided valuable inputs to the design of in-vehicle explanations for future AVs.

Although this study reveals valuable design recommendations for in-vehicle explanation systems, it has several limitations that require further investigation. First, to control the explanation contents as a confounding variable, I adopted the same information segments between the agents providing Level 2 and Level 3 SA and the other two types of agents who only provided one piece of information. In this case, the explanations supporting only Level 2 and Level 3 SA did not necessarily align with the drivers' information needs or consisted of complete information to reach sufficient support, which led to driver confusion and surprise. Future studies can continue exploring the appropriate approaches to support Level 2 or Level 3 information solely without losing explanation components, which will help identify the fine line between providing sufficient information without overloading participants. Second, although the message was designed to support drivers' SA, I did not measure SA in addition to a driver's driving performance and subjective evaluations. Specifically, Study 5 focused on exploring user needs through multiple measures at different timings. In addition, considering that post-explanation measures have already been administrated to understand drivers' perceptions of the explanations, additional measures

were not adopted that might extend the pause time and result in loss of immersion in driving simulation. Future studies could directly measure drivers' SA and validate the effects of providing explanations related to one or more levels of SA. Third, findings from two qualitative studies described in Chapter 4 indicate that explanation adaptability is not only limited to driver-automation function allocation, but also includes considering contextual factors and individual differences. Examples of contextual factors include urgency of the event, safety criticality of the event, and location of the event. Examples of individual differences include technology literacy, experiences with AVs, and driving experience. Future studies can also consider adapting the explanations based on contextual and individual factors. However, the mindset of providing explanations case by case is impossible, given that the list of events with explanation needs can be unlimited. Therefore, the ultimate goal of providing explanations is to develop a "function" that can generate explanations based on several inputs, such as the model-agnostic transparency in computer science. Finally, the longitudinal effects of explanations were not investigated in this single study. As pointed out in expert interviews, appropriate training and education are needed for drivers with future AVs. Good explanations can facilitate such training and education processes. In fact, during the experiments, some participants indicated that they got used to the agents in their third drive and did not want to hear the messages. This situation suggests that as drivers get familiar with the automation systems, their explanation needs decrease. Future studies can also explore how drivers' needs evolve as they get more experienced with the technology and how to customize explanations based on that.

CONCLUSION

The overarching goal of this dissertation was to explore the possibilities and appropriate design requirements of using intelligent agents to support drivers in conditionally automated driving. I explored user challenges in understanding and utilizing ADAS technologies and the potential solutions to address user confusion, misuse, and disuse through providing proper explanations. To achieve these goals, I first conducted two workshops to understand the design requirements for in-vehicle intelligent agents (IVIAs) and then selected two levels of embodiment conditions and speech styles to examine the effectiveness of different IVIA designs when supporting drivers' primary driving, which enabled me to understand the preferred IVIA form factors under conditionally automated driving. In Study 3, I strived to understand how existing users utilized their ADAS technologies and their challenges through contextual inquires, aiming to extract insights that can be transformed into conditionally automated driving conditions. In Study 4, potential solutions to address user challenges were explored through expert interviews. Findings from both Study 3 and Study 4 were consistent and merged into design recommendations for providing in-vehicle explanations. Finally, based on the findings from Studies 1-4 and existing theoretical frameworks, Study 5 focused on a selective aspect of developing explanations for conditionally automated vehicles – adapting explanations based on human-automation function allocation – in the format of IVIAs.

This dissertation has several key contributions. First, it contributes to the general design recommendations and guidelines for in-vehicle intelligent agent designs under three key levels of automation (Level 0, Level 3, and Level 5) as delineated in SAE International (2021). These recommendations not only support the subsequent research activities, but also provide valuable insights for including speech user interfaces as one of the natural language interfaces for future automated vehicles. Although developing a comprehensive list of guidance for IVIAs requires more than one study, the findings from my dissertation can point out directions for researchers and practitioners regarding factors to be considered when designing IVIAs. Second, this dissertation has synthesized design considerations for providing

explanations in automated vehicles (AVs) based on the inputs from both end users and subject matter experts. These design considerations not only provide guidelines and directions when designing explanations for AVs, but also pinpoint several traps that researchers and practitioners might fall into when promoting explainability of future AVs. Considering that these considerations were extracted based on actual user inputs and experts in both academia and industry, they are less subject to bias compared to those integrating only a limited type of stakeholders. Although these design considerations are specifically synthesized in the context of AVs, they are very likely to be transformed into other domains with AI-powered decision support systems. In addition, these design guidelines can be utilized to develop training and educational materials to reinforce drivers' understanding of future AVs. Third, this dissertation also emphasizes the potential incompatibility between the subjective ratings and driving behaviors. This incompatibility indicates that drivers' preferred system configuration might not yield better driving performance, which reveals the importance of balancing user satisfaction and performance. Such incompatibility serves as a reminder for both researchers and practitioners that only relying on subjective evaluations or focusing on improving user experiences might not be sufficient to deliver safer and more balanced solutions for future AVs. Finally, this dissertation has several theoretical contributions that indicate the benefit of providing explanations to support different stages of decision-making and prevent heuristic biases associated with automatic processing and response selection.

In summary, this dissertation not only contributes to design guidelines and evaluation methods to promote explainability of future AVs but also poses challenges in introducing explanations in advanced technologies. Findings from this dissertation are able to serve as the steppingstone to develop adaptive in-vehicle explanation systems that support drivers' accurate mental models while enhancing their user experience, encouraging appropriate usage of advanced technologies for overall traffic safety.

BIBLIOGRAPHY

Abraham, H., Reimer, B., & Mehler, B. (2018). Learning to Use In-Vehicle Technologies: Consumer Preferences and Effects on Understanding. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *62*(1), 1589–1593. https://doi.org/10.1177/1541931218621359

Atakishiyev, S., Salameh, M., Yao, H., & Goebel, R. (2023). *Explainable Artificial Intelligence for Autonomous Driving: A Comprehensive Overview and Field Guide for Future Research Directions* (arXiv:2112.11561). arXiv. http://arxiv.org/abs/2112.11561

Avetisyan, L., Ayoub, J., & Zhou, F. (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. https://doi.org/10.1016/j.trf.2022.07.010

Barredo Arrieta, A., Díaz-Rodríguez, N., Del Ser, J., Bennetot, A., Tabik, S., Barbado, A., Garcia, S., Gil-Lopez, S., Molina, D., Benjamins, R., Chatila, R., & Herrera, F. (2020). Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI. *Information Fusion*, *58*, 82–115. https://doi.org/10.1016/j.inffus.2019.12.012

Bartneck, C., Kulić, D., Croft, E., & Zoghbi, S. (2009). Measurement instruments for the anthropomorphism, animacy, likeability, perceived intelligence, and perceived safety of robots. *International Journal of Social Robotics*, *1*(1), 71–81. https://doi.org/10.1007/s12369-008-0001-3

Bengler, K., Rettenmaier, M., Fritz, N., & Feierle, A. (2020). From HMI to HMIs: Towards an HMI Framework for Automated Driving. *Information*, *11*(2), 61. https://doi.org/10.3390/info11020061

Bennett, K. B., & Flach, J. M. (2011a). Display Design: Building a Conceptual Base. In *Display and Interface Design* (pp. 142–146).

Bennett, K. B., & Flach, J. M. (2011b). Visual Attention and Form Perception. In *Display and Interface Design* (pp. 169–195). CRC Press.

Beringhoff, F., Greenyer, J., Roesener, C., & Tichy, M. (2022). Thirty-One Challenges in Testing Automated Vehicles: Interviews with Experts from Industry and Research. *2022 IEEE Intelligent Vehicles Symposium (IV)*, 360–366. https://doi.org/10.1109/IV51971.2022.9827097

Bhaskara, A., Skinner, M., & Loft, S. (2020). Agent transparency: A review of current theory and evidence. *IEEE Transactions on Human-Machine Systems*, *50*(3), 215–224. https://doi.org/10.1109/THMS.2020.2965529

Blattner, M., Sumikawa, D., & Greenberg, R. (1989). Earcons and Icons: Their Structure and Common Design Principles. *Human-Computer Interaction*, *4*(1), 11–44. https://doi.org/10.1207/s15327051hci0401_1

Bogner, A., & Menz, W. (2009). The Theory-Generating Expert Interview: Epistemological Interest, Forms of Knowledge, Interaction. In A. Bogner, B. Littig, & W. Menz (Eds.), *Interviewing Experts* (pp. 43–80). Palgrave Macmillan UK. https://doi.org/10.1057/9780230244276_3

Braun, M., Mainz, A., Chadowitz, R., Pfleging, B., & Alt, F. (2019). At Your Service: Designing Voice Assistant Personalities to Improve Automotive User Interfaces. *Conference on Human Factors in Computing Systems - Proceedings*, 1–11.

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, *3*(2), 77–101. https://doi.org/10.1191/1478088706qp063oa

Braun, V., & Clarke, V. (2012). Thematic analysis. In H. Cooper, P. M. Camic, D. L. Long, A. T. Panter, D. Rindskopf, & K. J. Sher (Eds.), *APA handbook of research methods in psychology, Vol 2: Research designs: Quantitative, qualitative, neuropsychological, and biological.* (pp. 57–71). American Psychological Association. https://doi.org/10.1037/13620-004

Brink, K. A., & Wellman, H. M. (2020). Robot Teachers for Children? Young Children Trust Robots Depending on Their Perceived Accuracy and Agency. *Developmental Psychology*, *56*(7), 1268–1277. https://doi.org/10.1037/dev0000884

Capallera, M., Angelini, L., Meteier, Q., Khaled, O. A., & Mugellini, E. (2023). Human-Vehicle Interaction to Support Driver's Situation Awareness in Automated Vehicles: A Systematic

Review. *IEEE Transactions on Intelligent Vehicles*, *8*(3), 2551–2567. https://doi.org/10.1109/TIV.2022.3200826

Carpinella, C. M., Wyman, A. B., Perez, M. A., & Stroessner, S. J. (2017). The robotic social attributes scale (RoSAS): Development and validation. *ACM/IEEE International Conference on Human-Robot Interaction*, *Part F1271*(March 2017), 254–262. https://doi.org/10.1145/2909824.3020208

Chen, J. Y. C., Barnes, M. J., Wright, J. L., Stowers, K., & Lakhmani, S. G. (2017). Situation awareness-based agent transparency for human-autonomy teaming effectiveness. *Micro- and Nanotechnology Sensors, Systems, and Applications IX*, *10194*(May 2017), 101941V. https://doi.org/10.1117/12.2263194

Choe, M., & Jeon, M. (2023). "I See You": Comparing the Effects of Affective Empathy and Cognitive Empathy on Drivers' Affective States and Driving Behavior in Frustrating Driving Contexts. *Adjunct Proceedings of the 15th International Conference on Automotive User Interfaces and Interactive Vehicular Applications*, 121–126. https://doi.org/10.1145/3581961.3609879

Cohen, J., Cohen, P., West, S. G., & Aiken, L. S. (2013). *Applied Multiple Regression/Correlation Analysis for the Behavioral Sciences*. Routledge. https://doi.org/10.4324/9780203774441

Colley, M., Britten, J., Demharter, S., Hisir, T., & Rukzio, E. (2022). Feedback Strategies for Crowded Intersections in Automated Traffic—A Desirable Future? *Proceedings of the 14th International Conference on Automotive User Interfaces and Interactive Vehicular Applications*, 243–252. https://doi.org/10.1145/3543174.3545255

Colley, M., Krauss, S., Lanzer, M., & Rukzio, E. (2021). How Should Automated Vehicles Communicate Critical Situations?: A Comparative Analysis of Visualization Concepts. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, *5*(3), 1–23. https://doi.org/10.1145/3478111

Dandekar, A., Mathis, L.-A., Berger, M., & Pfleging, B. (2022). How to Display Vehicle Information to Users of Automated Vehicles When Conducting Non-Driving-Related Activities. *Proceedings of the ACM on Human-Computer Interaction*, *6*(MHCI), 1–22. https://doi.org/10.1145/3546741

Davis, F. (1986). *A technology acceptance model for empirically testing new end-user information systems: Theory and results.* [Doctoral dissertation]. MIT Sloan School of Management.

De Groot, J. H. B., Semin, G. R., & Smeets, M. A. M. (2014). I can see, hear, and smell your fear: Comparing olfactory and audiovisual media in fear communication. *Journal of Experimental Psychology: General*, *143*(2), 825–834. https://doi.org/10.1037/a0033731

De Winter, J. C. F., Happee, R., Martens, M. H., & Stanton, N. A. (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. https://doi.org/10.1016/j.trf.2014.06.016

Dingus, T. A., Guo, F., Lee, S., Antin, J. F., Perez, M., Buchanan-King, M., & Hankey, J. (2016). Driver crash risk factors and prevalence evaluation using naturalistic driving data. *Proceedings of the National Academy of Sciences of the United States of America*, *113*(10), 2636–2641. https://doi.org/10.1073/pnas.1513271113

Dingus, T. A., Hankey, J. M., Antin, J. F., Lee, S. E., Eichelberger, L., Stulce, K., McGraw, D., Perez, M., Stowe, L., Strategic Highway Research Program Safety Focus Area, & Transportation Research Board. (2014). *Naturalistic Driving Study: Technical Coordination and Quality Control* (p. 22362). Transportation Research Board. https://doi.org/10.17226/22362

Dixon, L., Schneider, N., Usai, M., Herzberger, N. D., Flemisch, F. O., & Baumann, M. (2023). Exploring Driver Responses to Authoritative Control Interventions in Highly Automated Driving. *Proceedings of the 15th International Conference on Automotive User Interfaces and Interactive Vehicular Applications*, 145–155. https://doi.org/10.1145/3580585.3607159

Dogan, E., Rahal, M.-C., Deborne, R., Delhomme, P., Kemeny, A., & Perrin, J. (2017). Transition of control in a partially automated vehicle: Effects of anticipation and non-driving-related task involvement. *Transportation Research Part F: Traffic Psychology and Behaviour*, *46*, 205–215. https://doi.org/10.1016/j.trf.2017.01.012

Dong, J., Lawson, E., Olsen, J., & Jeon, M. (2020). Female voice agents in fully autonomous vehicles are not only more likeable and comfortable, but also more competent. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *64*(1), 1033–1037. https://doi.org/10.1177/1071181320641248

Dozza, M. (2013). What factors influence drivers' response time for evasive maneuvers in real traffic? *Accident Analysis & Prevention*, *58*, 299–308. https://doi.org/10.1016/j.aap.2012.06.003

Du, N., Haspiel, J., Zhang, Q., Tilbury, D., Pradhan, A. K., Yang, X. J., & Robert, L. P. (2019). 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. https://doi.org/10.1016/j.trc.2019.05.025

Du, N., Zhou, F., Tilbury, D., Robert, L. P., & Yang, X. J. (2021). Designing alert systems in takeover transitions: The effects of display information and modality. *Proceedings - 13th International ACM Conference on Automotive User Interfaces and Interactive Vehicular Applications, AutomotiveUI 2021*, 173–180. https://doi.org/10.1145/3409118.3475155

Dunn, N. J., Dingus, T. A., Soccolich, S., & Horrey, W. J. (2021). Investigating the impact of driving automation systems on distracted driving behaviors. *Accident Analysis & Prevention*, *156*, 106152. https://doi.org/10.1016/j.aap.2021.106152

Eby, D. W., Molnar, L. J., Zakrajsek, J. S., Ryan, L. H., Zanier, N., Louis, R. M. St., Stanciu, S. C., LeBlanc, D., Kostyniuk, L. P., Smith, J., Yung, R., Nyquist, L., DiGuiseppi, C., Li, G., Mielenz, T. J., & Strogatz, D. (2018). Prevalence, attitudes, and knowledge of in-vehicle technologies and vehicle adaptations among older drivers. *Accident Analysis & Prevention*, *113*, 54–62. https://doi.org/10.1016/j.aap.2018.01.022

Endsley, M. R. (1988). Situation awareness global assessment technique (SAGAT). *Proceedings of the IEEE 1988 National Aerospace and Electronics Conference*, 789–795. https://doi.org/10.1109/NAECON.1988.195097

Endsley, M. R. (1995). Toward a Theory of Situation Awareness in Dynamic Systems. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *37*(1), 32–64. https://doi.org/10.1518/001872095779049543

Endsley, M. R. (2016). *Designing for Situation Awareness: An Approach to User-Centered Design* (2nd ed.). CRC Press. https://doi.org/10.1201/b11371

Endsley, M. R. (2021). *Situation Awareness Measurement: How to Measure Situation Awareness in Individuals and Teams*. Human Factors and Ergonomics Society.

Endsley, M. R., & Garland, D. J. (2000). *Situation Awareness Analysis and Measurement* (M. R. Endsley & D. J. Garland, Eds.). CRC Press. https://doi.org/10.1201/b12461

Feng, F., Bao, S., Sayer, J. R., Flannagan, C., Manser, M., & Wunderlich, R. (2017). Can vehicle longitudinal jerk be used to identify aggressive drivers? An examination using naturalistic driving data. *Accident Analysis & Prevention*, *104*, 125–136. https://doi.org/10.1016/j.aap.2017.04.012

Flick, U. (2023). *An introduction to qualitative research* (7th edition). SAGE.

Forster, Y., Naujoks, F., & Neukum, A. (2017). Increasing anthropomorphism and trust in automated driving functions by adding speech output. *2017 IEEE Intelligent Vehicles Symposium (IV)*, *2*(Iv), 365–372. https://doi.org/10.1109/IVS.2017.7995746

Gable, T. M., & Walker, B. N. (2013). Georgia Tech Simulator Sickness Screening Protocol. *Georgia Tech School of Psychology Tech Report GT-PSYC-TR-2013-01*, 1–16.

Gaffar, A., & Kouchak, S. M. (2018). *Using simplified grammar for voice commands to decrease driver distraction. March*.

Geiser, G. (1985). Man machine interaction in vehicles. *Atz*, *87*(74–77), 56.

Gellatly, A. W., Hansen, C., Highstrom, M., & Weiss, J. P. (2010). Journey: General Motors' move to incorporate contextual design into its next generation of automotive HMI designs. *Proceedings of the 2nd International Conference on Automotive User Interfaces and Interactive Vehicular Applications*, 156–161. https://doi.org/10.1145/1969773.1969802

Ghizlene, B., Zoulikha, M., & Pomares, H. (2019). An Efficient Framework to Detect and Avoid Driver Sleepiness Based on YOLO with Haar Cascades and an Intelligent Agent. In *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics): Vol. 11507 LNCS*. Springer International Publishing. https://doi.org/10.1007/978-3-030-20518-8_58

Gläser, J., & Laudel, G. (2009). On Interviewing "Good" and "Bad" Experts. In A. Bogner, B. Littig, & W. Menz (Eds.), *Interviewing Experts* (pp. 117–137). Palgrave Macmillan UK. https://doi.org/10.1057/9780230244276_6

Gold, C., Körber, M., Lechner, D., & Bengler, K. (2016). Taking Over Control From Highly Automated Vehicles in Complex Traffic Situations: The Role of Traffic Density. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *58*(4), 642–652. https://doi.org/10.1177/0018720816634226

Graefe, J., Paden, S., Engelhardt, D., & Bengler, K. (2022). Human Centered Explainability for Intelligent Vehicles – A User Study. *Proceedings of the 14th International Conference on Automotive User Interfaces and Interactive Vehicular Applications*, 297–306. https://doi.org/10.1145/3543174.3546846

Gruden, T., Tomažič, S., Sodnik, J., & Jakus, G. (2022). A user study of directional tactile and auditory user interfaces for take-over requests in conditionally automated vehicles. *Accident Analysis & Prevention*, *174*, 106766. https://doi.org/10.1016/j.aap.2022.106766

Gunning, D., & Aha, D. W. (2019). DARPA's Explainable Artificial Intelligence Program. *AI Magazine*, *40*(2), 44–58. https://doi.org/10.1609/aimag.v40i2.2850

Hale, A. R., Stoop, J., & Hommels, J. (1990). Human error models as predictors of accident scenarios for designers in road transport systems. *Ergonomics*, *33*(10–11), 1377–1387. https://doi.org/10.1080/00140139008925339

Hartwich, F., Hollander, C., Johannmeyer, D., & Krems, J. F. (2021). Improving Passenger Experience and Trust in Automated Vehicles Through User-Adaptive HMIs: "The More the Better" Does

Not Apply to Everyone. *Frontiers in Human Dynamics*, *3*, 669030. https://doi.org/10.3389/fhumd.2021.669030

Hester, M., Lee, K., & Dyre, B. P. (2017). "Driver take over": A preliminary exploration of driver trust and performance in autonomous vehicles. *Proceedings of the Human Factors and Ergonomics Society*, *2017-Octob*, 1969–1973. https://doi.org/10.1177/1541931213601971

Hoffman, R. R., Mueller, S. T., Klein, G., & Litman, J. (2018). *Metrics for Explainable AI: Challenges and Prospects*. 1–50. https://doi.org/10.48550/arXiv.1812.04608

Hoffman, R. R., Mueller, S. T., Klein, G., & Litman, J. (2023). Measures for explainable AI: Explanation goodness, user satisfaction, mental models, curiosity, trust, and human-AI performance. *Frontiers in Computer Science*, *5*, 1096257. https://doi.org/10.3389/fcomp.2023.1096257

Holthausen, B. E., Wintersberger, P., Walker, B. N., & Riener, A. (2020). Situational Trust Scale for Automated Driving (STS-AD): Development and Initial Validation. *Proceedings - 12th International ACM Conference on Automotive User Interfaces and Interactive Vehicular Applications, AutomotiveUI 2020*, 40–47. https://doi.org/10.1145/3409120.3410637

Holtzblatt, K., & Beyer, H. (2017a). *Contextual Design: Design for Life* (Second). San Francisco: Elsevier Science.

Holtzblatt, K., & Beyer, H. (2017b). Principles of Contextual Inquiry. In *Contextual Design* (pp. 43–80). Elsevier. https://doi.org/10.1016/B978-0-12-800894-2.00003-X

Holtzblatt, K., & Beyer, H. (2017c). The affinity diagram. In K. Holtzblatt & H. Beyer (Eds.), *Contextual design* (pp. 127–146). Elsevier. https://doi.org/10.1016/b978-0-12-800894-2.00006-5

Holzinger, A., Saranti, A., Molnar, C., Biecek, P., & Samek, W. (2022). Explainable AI Methods—A Brief Overview. In A. Holzinger, R. Goebel, R. Fong, T. Moon, K.-R. Müller, & W. Samek (Eds.), *xxAI - Beyond Explainable AI* (Vol. 13200, pp. 13–38). Springer International Publishing. https://doi.org/10.1007/978-3-031-04083-2_2

Hone, K. S., & Graham, R. (2000). Towards a tool for the subjective assessment of speech system interfaces (SASSI). *Natural Language Engineering*, *6*(3–4), 287–291. https://doi.org/10.1017/s1351324900002497

Hong, S., Maeng, J., Kim, H. J., & Yang, J. H. (2022). Development of Warning Methods for Planned and Unplanned Takeover Requests in a Simulated Automated Driving Vehicle. *Proceedings of the 14th International Conference on Automotive User Interfaces and Interactive Vehicular Applications*, 65–74. https://doi.org/10.1145/3543174.3545999

Hosseini, S. M. F., Lettinga, D., Vasey, E., Zheng, Z., Jeon, M., Park, C. H., & Howard, A. M. (2017). Both look and feel matter: Essential factors for robotic companionship. *RO-MAN 2017 - 26th IEEE International Symposium on Robot and Human Interactive Communication*, *2017-Janua*, 150–155. https://doi.org/10.1109/ROMAN.2017.8172294

Huang, G., & Pitts, B. J. (2022). To Inform or to Instruct? An Evaluation of Meaningful Vibrotactile Patterns to Support Automated Vehicle Takeover Performance. *IEEE Transactions on Human-Machine Systems*, 1–10. https://doi.org/10.1109/THMS.2022.3205880

Hubbard, S. M. L. (2018). Automated Vehicle Legislative Issues. *Transportation Research Record: Journal of the Transportation Research Board*, *2672*(7), 1–13. https://doi.org/10.1177/0361198118774155

Islam, S. R., Eberle, W., Ghafoor, S. K., & Ahmed, M. (2021). *Explainable Artificial Intelligence Approaches: A Survey* (arXiv:2101.09429). arXiv. http://arxiv.org/abs/2101.09429

Jeon, M. (2015). Towards affect-integrated driving behaviour research. *Theoretical Issues in Ergonomics Science*, *16*(6), 553–585. https://doi.org/10.1080/1463922X.2015.1067934

Jeon, M. (2016). Don't Cry While You're Driving: Sad Driving Is as Bad as Angry Driving. *International Journal of Human-Computer Interaction*, *32*(10), 777–790. https://doi.org/10.1080/10447318.2016.1198524

Jeon, M. (2019). Multimodal Displays for Take-over in Level 3 Automated Vehicles while Playing a Game. *Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems*, 1–6. https://doi.org/10.1145/3290607.3313056

Jeon, M., Lautala, P. T., Nadri, C., & Nelson, D. N. (2022). *In-Vehicle Auditory Alerts Literature Review* (Issue February). https://rosap.ntl.bts.gov/view/dot/60401

Jeon, M., Nadri, C., Wang, M., & Dam, A. (2024). Chapter 9 Automotive User Interfaces. In C. Stephanidis & G. Salvendy (Eds.), *Human-Computer Interaction in Intelligent Environments* (First edition). CRC Press.

Jeon, M., & Walker, B. N. (2011). What to detect? Analyzing factor structures of affect in driving contexts for an emotion detection and regulation system. *Proceedings of the Human Factors and Ergonomics Society*, 1889–1893. https://doi.org/10.1177/1071181311551393

Jeon, M., Walker, B. N., & Gable, T. M. (2015). The effects of social interactions with in-vehicle agents on a driver's anger level, driving performance, situation awareness, and perceived workload. *Applied Ergonomics*, *50*, 185–199. https://doi.org/10.1016/j.apergo.2015.03.015

Jeon, M., Walker, B. N., & Yim, J. B. (2014). Effects of specific emotions on subjective judgment, driving performance, and perceived workload. *Transportation Research Part F: Traffic Psychology and Behaviour*, *24*, 197–209. https://doi.org/10.1016/j.trf.2014.04.003

Jian, J.-Y., Bisantz, A. M., & Drury, C. G. (2000). Foundations for an empirically determined scale of trust in automated systems. *International Journal of Cognitive Ergonomics*, *4*(1), 53–71. https://doi.org/10.1207/S15327566IJCE0401_04

Johnsson, I.-M., Nass, C., Harris, H., & Takayama, L. (2005). Matching In-Car Voice with Driver State: Impact on Attitude and Driving Performance. *Driving Assessment 2005 : Proceedings of the 3rd International Driving Symposium on Human Factors in Driver Assessment, Training, and Vehicle Design*, 173–180. https://doi.org/10.17077/drivingassessment.1158

Josephson, J. R., & Josephson, S. G. (1996). *Abductive inference: Computation, philosophy, technology*. Cambridge University Press.

Kim, H., Martin, S., Tawari, A., Misu, T., & Gabbard, J. L. (2020). Toward Real-Time Estimation of Driver Situation Awareness: An Eye-tracking Approach based on Moving Objects of Interest. *2020 IEEE Intelligent Vehicles Symposium (IV)*, 1035–1041. https://doi.org/10.1109/IV47402.2020.9304770

Koo, J., Kwac, J., Ju, W., Steinert, M., Leifer, L., & Nass, C. (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*, *9*(4), 269–275. https://doi.org/10.1007/s12008-014-0227-2

Kraus, J., Nothdurft, F., Hock, P., Scholz, D., Minker, W., & Baumann, M. (2016). Human after all: Effects of mere presence and social interaction of a humanoid robot as a co-driver in automated driving. *AutomotiveUI 2016 - 8th International Conference on Automotive User Interfaces and Interactive Vehicular Applications, Adjunct Proceedings*, 129–134. https://doi.org/10.1145/3004323.3004338

Kraus, J., Scholz, D., Stiegemeier, D., & Baumann, M. (2020). The More You Know: Trust Dynamics and Calibration in Highly Automated Driving and the Effects of Take-Overs, System Malfunction, and System Transparency. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *62*(5), 718–736. https://doi.org/10.1177/0018720819853686

Kraus, J., Sturn, J., Reiser, J. E., & Baumann, M. (2015). Anthropomorphic agents, transparent automation and driver personality: Towards an integrative multi-level model of determinants for effective driver-vehicle cooperation in highly automated vehicles. *Adjunct Proceedings of the 7th International Conference on Automotive User Interfaces and Interactive Vehicular Applications - AutomotiveUI '15*, 8–13. https://doi.org/10.1145/2809730.2809738

Krome, S., Walz, S. P., & Greuter, S. (2016). Contextual Inquiry of Future Commuting in Autonomous Cars. *Proceedings of the 2016 CHI Conference Extended Abstracts on Human Factors in Computing Systems*, 3122–3128. https://doi.org/10.1145/2851581.2892336

Kundinger, T., Wintersberger, P., & Riener, A. (2019). (Over)Trust in Automated Driving: The Sleeping Pill of Tomorrow? *Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems*, 1–6. https://doi.org/10.1145/3290607.3312869

Large, D. R., Harrington, K., Burnett, G., Luton, J., Thomas, P., & Bennett, P. (2019). To please in a pod: Employing an anthropomorphic agent-interlocutor to enhance trust and user experience in an autonomous, self-driving vehicle. *Proceedings - 11th International ACM Conference on Automotive User Interfaces and Interactive Vehicular Applications, AutomotiveUI 2019*, 49–59. https://doi.org/10.1145/3342197.3344545

Lau, M., Wilbrink, M., Dodiya, J., & Oehl, M. (2020). Users' Internal HMI Information Requirements for Highly Automated Driving. In C. Stephanidis, M. Antona, & S. Ntoa (Eds.), *HCI International 2020 – Late Breaking Posters* (Vol. 1294, pp. 585–592). Springer International Publishing. https://doi.org/10.1007/978-3-030-60703-6_75

Lavan, N., Burton, A. M., Scott, S. K., & McGettigan, C. (2019). Flexible voices: Identity perception from variable vocal signals. *Psychonomic Bulletin and Review*, *26*(1), 90–102. https://doi.org/10.3758/s13423-018-1497-7

Lee, J. D., & See, K. A. (2004). Trust in automation: Designing for appropriate reliance. *Human Factors*, *46*(1), 50–80. https://doi.org/10.1518/hfes.46.1.50_30392

Lee, J. D., Wickens, C. D., Liu, Y., & Boyle, L. N. (2017a). Chapter 7: Decision Making and Macrocognition. In *Designing for people: An introduction to human factors engineering* (3rd ed., pp. 203–238). CreateSpace.

Lee, J. D., Wickens, C. D., Liu, Y., & Boyle, L. N. (2017b). Chapter 8: Displays. In *Designing for people: An introduction to human factors engineering* (3rd ed., pp. 246–252). CreateSpace.

Lee, J. D., Wickens, C. D., Liu, Y., & Boyle, L. N. (2017c). *Designing for people: An introduction to human factors engineering*. CreateSpace.

Lee, S. C., & Jeon, M. (2022). A Systematic Review of Functions and Design Features of In-Vehicle Agents. *International Journal of Human-Computer Studies*, 102864. https://doi.org/10.1016/J.IJHCS.2022.102864

Lee, S. C., Jeong, S., Wang, M., Hock, P., Baumann, M., & Jeon, M. (2021). To Go or Not to Go? That is the Question When In-Vehicle Agents Argue with Each Other. *Adjunct Proceedings - 13th International ACM Conference on Automotive User Interfaces and Interactive Vehicular Applications, AutomotiveUI 2021*, 223–224. https://doi.org/10.1145/3473682.3481876

Lee, S. C., Nadri, C., Sanghavi, H., & Jeon, M. (2022). Eliciting User Needs and Design Requirements for User Experience in Fully Automated Vehicles. *International Journal of Human–Computer Interaction*, *38*(3), 227–239. https://doi.org/10.1080/10447318.2021.1937875

Lewis, B. A., Eisert, J. L., & Baldwin, C. L. (2018). Validation of Essential Acoustic Parameters for Highly Urgent In-Vehicle Collision Warnings. *Human Factors*, *60*(2), 248–261. https://doi.org/10.1177/0018720817742114

Li, S., Blythe, P., Guo, W., Namdeo, A., Edwards, S., Goodman, P., & Hill, G. (2019). Evaluation of the effects of age-friendly human-machine interfaces on the driver's takeover performance in highly automated vehicles. *Transportation Research Part F: Traffic Psychology and Behaviour*, *67*, 78–100. https://doi.org/10.1016/j.trf.2019.10.009

Löcken, A., Frison, A.-K., Fahn, V., Kreppold, D., Götz, M., & Riener, A. (2020). Increasing User Experience and Trust in Automated Vehicles via an Ambient Light Display. *22nd International Conference on Human-Computer Interaction with Mobile Devices and Services*, 1–10. https://doi.org/10.1145/3379503.3403567

Lubkowski, S. D., Lewis, B. A., Gawron, V. J., Gaydos, T. L., Campbell, K. C., Kirkpatrick, S. A., Reagan, I. J., & Cicchino, J. B. (2021). Driver trust in and training for advanced driver assistance systems in Real-World driving. *Transportation Research Part F: Traffic Psychology and Behaviour*, *81*, 540–556. https://doi.org/10.1016/j.trf.2021.07.003

Lyons, J. B. (2013). Being transparent about transparency: A model for human-robot interaction. *AAAI Spring Symposium - Technical Report*, *SS-13-07*, 48–53.

Mahajan, K., Large, D. R., Burnett, G., & Velaga, N. R. (2021a). Exploring the benefits of conversing with a digital voice assistant during automated driving: A parametric duration model of takeover time. *Transportation Research Part F: Traffic Psychology and Behaviour*, *80*, 104–126. https://doi.org/10.1016/j.trf.2021.03.012

Mahajan, K., Large, D. R., Burnett, G., & Velaga, N. R. (2021b). Exploring the effectiveness of a digital voice assistant to maintain driver alertness in partially automated vehicles. *Traffic Injury Prevention*, *22*(5), 378–383. https://doi.org/10.1080/15389588.2021.1904138

Manger, C., Peintner, J., Hoffmann, M., Probst, M., Wennmacher, R., & Riener, A. (2023). Providing Explainability in Safety-Critical Automated Driving Situations through Augmented Reality Windshield HMIs. *Adjunct Proceedings of the 15th International Conference on Automotive User Interfaces and Interactive Vehicular Applications*, 174–179. https://doi.org/10.1145/3581961.3609874

Mason, J., Carney, C., Gaspar, J. G., Kim, W., Romo, A., & Horrey, W. J. (2023). *Mapping Comprehension of ADAS across Different Road Users*. AAA Foundation for Traffic Safety.

Mathias, S. R., & von Kriegstein, K. (2019). *Voice Processing and Voice-Identity Recognition*. https://doi.org/10.1007/978-3-030-14832-4_7

McDonald, A. D., Alambeigi, H., Engström, J., Markkula, G., Vogelpohl, T., Dunne, J., & Yuma, N. (2019). Toward Computational Simulations of Behavior During Automated Driving Takeovers: A Review of the Empirical and Modeling Literatures. *Human Factors*, *61*(4), 642–688. https://doi.org/10.1177/0018720819829572

Mehrotra, S., Wang, M., Wong, N., Parker, J., Robters, S. C., Kim, W., Romo, A., & Horrey, W. J. (2022). *Human-Machine Interfaces and Vehicle Automation: A Review of the Literature and Recommendations for System Design, Feedback, and Alerts*.

Merritt, S. M., Heimbaugh, H., LaChapell, J., & Lee, D. (2013). I Trust It, but I Don't Know Why: Effects of Implicit Attitudes Toward Automation on Trust in an Automated System. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *55*(3), 520–534. https://doi.org/10.1177/0018720812465081

Merritt, S. M., & Ilgen, D. R. (2008). Not All Trust Is Created Equal: Dispositional and History-Based Trust in Human-Automation Interactions. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *50*(2), 194–210. https://doi.org/10.1518/001872008X288574

Meschtscherjakov, A., Wilfinger, D., Gridling, N., Neureiter, K., & Tscheligi, M. (2011). Capture the car!: Qualitative in-situ methods to grasp the automotive context. *Proceedings of the 3rd International Conference on Automotive User Interfaces and Interactive Vehicular Applications*, 105–112. https://doi.org/10.1145/2381416.2381434

Meuser, M., & Nagel, U. (2009). The Expert Interview and Changes in Knowledge Production. In A. Bogner, B. Littig, & W. Menz (Eds.), *Interviewing Experts* (pp. 17–42). Palgrave Macmillan UK. https://doi.org/10.1057/9780230244276_2

Michon, J. A. (1985). A Critical View of Driver Behavior Models: What Do We Know, What Should We Do? In *Human Behavior and Traffic Safety* (pp. 485–524). Springer US. https://doi.org/10.1007/978-1-4613-2173-6_19

Miller, T. (2019). Explanation in artificial intelligence: Insights from the social sciences. *Artificial Intelligence*, *267*, 1–38. https://doi.org/10.1016/j.artint.2018.07.007

Miller, T. (2023, March 10). *Explainable AI is Dead, Long Live Explainable AI! Hypothesis-driven decision support*. http://arxiv.org/abs/2302.12389

Monticello, M. (2023). Ford's BlueCruise Ousts GM's Super Cruise as CR's Top-Rated Active Driving Assistance System. *Consumerreports.Org*. https://www.consumerreports.org/cars/car-safety/active-driving-assistance-systems-review-a2103632203/

Muir, B. M., & Moray, N. (1996). Trust in automation. Part II. Experimental studies of trust and human intervention in a process control simulation. *Ergonomics*, *39*(3), 429–460. https://doi.org/10.1080/00140139608964474

Nass, C., Jonsson, I. M., Harris, H., Reaves, B., Endo, J., Brave, S., & Takayama, L. (2005). Improving automotive safety by pairing driver emotion and car voice emotion. *Conference on Human Factors in Computing Systems - Proceedings*, 1973–1976. https://doi.org/10.1145/1056808.1057070

Nass, C., Steuer, J., & Tauber, E. R. (1994). Computers are social actors. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems Celebrating Interdependence - CHI '94*, 72–78. https://doi.org/10.1145/191666.191703

Naujoks, F., Wiedemann, K., & Schömig, N. (2017). The Importance of Interruption Management for Usefulness and Acceptance of Automated Driving. *Proceedings of the 9th International Conference on Automotive User Interfaces and Interactive Vehicular Applications*, 254–263. https://doi.org/10.1145/3122986.3123000

Nees, M. A., Helbein, B., & Porter, A. (2016). Speech Auditory Alerts Promote Memory for Alerted Events in a Video-Simulated Self-Driving Car Ride. *Human Factors*, *58*(3), 416–426. https://doi.org/10.1177/0018720816629279

Noble, A. M. (2020). *Behavioral Adaptation to Driving Automation Systems: Guidance for Consumer Education* [Virginia Tech]. http://hdl.handle.net/10919/105207

Noble, A. M., Klauer, S. G., Doerzaph, Z. R., & Manser, M. P. (2019). Driver Training for Automated Vehicle Technology – Knowledge, Behaviors, and Perceived Familiarity. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *63*(1), 2110–2114. https://doi.org/10.1177/1071181319631249

Novakazi, F., Johansson, M., Strömberg, H., & Karlsson, M. (2021). Levels of What?Investigating Drivers' Understanding of Different Levels of Automation in Vehicles. *Journal of Cognitive*

Engineering and Decision Making, *15*(2–3), 116–132. https://doi.org/10.1177/15553434211009024

Othman, M. R., Zhong Zhang, Takashi Imamura, & Tetsuo Miyake. (2008). A study of analysis method for driver features extraction. *2008 IEEE International Conference on Systems, Man and Cybernetics*, 1501–1505. https://doi.org/10.1109/ICSMC.2008.4811498

Parasuraman, R., & Manzey, D. H. (2010). Complacency and bias in human use of automation: An attentional integration. *Human Factors*, *52*(3), 381–410. https://doi.org/10.1177/0018720810376055

Parasuraman, R., & Riley, V. (1997). Humans and Automation: Use, Misuse, Disuse, Abuse. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *39*(2), 230–253. https://doi.org/10.1518/001872097778543886

Parasuraman, R., Sheridan, T. B., & Wickens, C. D. (2000). A model for types and levels of human interaction with automation. *IEEE Transactions on Systems, Man, and Cybernetics - Part A: Systems and Humans*, *30*(3), 286–297. https://doi.org/10.1109/3468.844354

Park, S., Xing, Y., Akash, K., Misu, T., & Boyle, L. N. (2022). The Impact of Environmental Complexity on Drivers' Situation Awareness. *Proceedings of the 14th International Conference on Automotive User Interfaces and Interactive Vehicular Applications*, 131–138. https://doi.org/10.1145/3543174.3546831

Parreira, M. T., & Gillet, S. (2022). *Design Implications for Effective Robot Gaze Behaviors in Multiparty Interactions*. *2017*, 976–980.

Pauzié, A. (2008). A method to assess the driver mental workload: The driving activity load index (DALI). *IET Intelligent Transport Systems*, *2*(4), 315. https://doi.org/10.1049/iet-its:20080023

Peintner, J. B., Manger, C., & Riener, A. (2022). "Can you rely on me?" Evaluating a Confidence HMI for Cooperative, Automated Driving. *Proceedings of the 14th International Conference on Automotive User Interfaces and Interactive Vehicular Applications*, 340–348. https://doi.org/10.1145/3543174.3546976

Petermeijer, S., De Winter, J. C. F., & Bengler, K. J. (2016). Vibrotactile Displays: A Survey With a View on Highly Automated Driving. *IEEE Transactions on Intelligent Transportation Systems*, *17*(4), 897–907. https://doi.org/10.1109/TITS.2015.2494873

Petermeijer, S., Doubek, F., & De Winter, J. (2017). Driver response times to auditory, visual, and tactile take-over requests: A simulator study with 101 participants. *2017 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, 1505–1510. https://doi.org/10.1109/SMC.2017.8122827

Pettersson, I., & Ju, W. (2017). Design Techniques for Exploring Automotive Interaction in the Drive towards Automation. *Proceedings of the 2017 Conference on Designing Interactive Systems*, 147–160. https://doi.org/10.1145/3064663.3064666

Phillips, P. J., Hahn, C. A., Fontana, P. C., Yates, A. N., Greene, K., Broniatowski, D. A., & Przybocki, M. A. (2021). *Four principles of explainable artificial intelligence* (NIST IR 8312; p. NIST IR 8312). National Institute of Standards and Technology (U.S.). https://doi.org/10.6028/NIST.IR.8312

Politis, I., Brewster, S., & Pollick, F. (2015). *Language-based multimodal displays for the handover of control in autonomous cars*. *c*, 3–10. https://doi.org/10.1145/2799250.2799262

Premstaller, M., Kotsios, H., & Wintersberger, P. (2023). Embodied Conversational Agent Teams for Trust Calibration in Automated Vehicles. *Adjunct Proceedings of the 15th International Conference on Automotive User Interfaces and Interactive Vehicular Applications*, 71–76. https://doi.org/10.1145/3581961.3609890

Ranney, T. A. (1994). Models of driving behavior: A review of their evolution. *Accident Analysis & Prevention*, *26*(6), 733–750. https://doi.org/10.1016/0001-4575(94)90051-5

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*, *SMC-13*(3), 257–266. https://doi.org/10.1109/TSMC.1983.6313160

Rheu, M., Shin, J. Y., Peng, W., & Huh-Yoo, J. (2020). Systematic Review: Trust-Building Factors and Implications for Conversational Agent Design. *International Journal of Human-Computer Interaction*, *00*(00), 1–16. https://doi.org/10.1080/10447318.2020.1807710

Richie, E., Offer-Westort, T., Shankar, R., & Jeon, M. (2018). Auditory Displays for Take-Over in Semi-automated Vehicles. In V. G. Duffy (Ed.), *Digital Human Modeling. Applications in Health, Safety, Ergonomics, and Risk Management* (Vol. 10917, pp. 623–634). Springer International Publishing. https://doi.org/10.1007/978-3-319-91397-1_51

Roche, F., & Brandenburg, S. (2020). Should the Urgency of Visual-Tactile Takeover Requests Match the Criticality of Takeover Situations. *IEEE Transactions on Intelligent Vehicles*, *5*(2), 306–313. https://doi.org/10.1109/TIV.2019.2955906

Roesler, E., Manzey, D., & Onnasch, L. (2021). A meta-Analysis on the effectiveness of anthropomorphism in human-robot interaction. *Science Robotics*, *6*(58). https://doi.org/10.1126/scirobotics.abj5425

Roscher, R., Bohn, B., Duarte, M. F., & Garcke, J. (2020). Explainable Machine Learning for Scientific Insights and Discoveries. *IEEE Access*, *8*, 42200–42216. https://doi.org/10.1109/ACCESS.2020.2976199

SAE International. (2021). Taxonomy and Definitions for Terms Related to Driving Automation Systems for On-Road Motor Vehicles. In *J3016_202104*.

SAE International. (2023). *Operational Definitions of Driving Performance Measures and Statistics*. SAE International. https://doi.org/10.4271/J2944_202302

Samrose, S., Anbarasu, K., Joshi, A., & Mishra, T. (2020). Mitigating Boredom Using An Empathetic Conversational Agent. *Proceedings of the 20th ACM International Conference on Intelligent Virtual Agents, IVA 2020*. https://doi.org/10.1145/3383652.3423905

Sanghavi, H., Jeon, M., Nadri, C., Ko, S., Sodnik, J., & Stojmenova, K. (2021). Multimodal Takeover Request Displays for Semi-automated Vehicles: Focused on Spatiality and Lead Time. In H.

Krömker (Ed.), *HCI in Mobility, Transport, and Automotive Systems* (Vol. 12791, pp. 315–334). Springer International Publishing. https://doi.org/10.1007/978-3-030-78358-7_22

Sanghavi, H., Zhang, Y., & Jeon, M. (2020). Effects of Anger and Display Urgency on Takeover Performance in Semi-automated Vehicles. *Proceedings - 12th International ACM Conference on Automotive User Interfaces and Interactive Vehicular Applications, AutomotiveUI 2020*, 48–56. https://doi.org/10.1145/3409120.3410664

Sarigul, B., Saltik, I., Hokelek, B., & Urgen, B. A. (2020). Does the Appearance of an Agent Affect How We Perceive his/her Voice? *Companion of the 2020 ACM/IEEE International Conference on Human-Robot Interaction*, 430–432. https://doi.org/10.1145/3371382.3378302

Schneider, T., Ghellal, S., Love, S., & Gerlicher, A. R. S. (2021). Increasing the User Experience in Autonomous Driving through different Feedback Modalities. *26th International Conference on Intelligent User Interfaces*, 7–10. https://doi.org/10.1145/3397481.3450687

Schuß, M., Manger, C., Löcken, A., Riener, A., & Riener, A. 2022. (2022). You'll Never Ride Alone: Insights into Women's Security Needs in Shared Automated Vehicles. *Proceedings of the 14th International Conference on Automotive User Interfaces and Interactive Vehicular Applications*, 13–23. https://doi.org/10.1145/3543174.3546848

Sharp, H., Rogers, Y., & Preece, J. (2019). Chapter 2: The process of interaction design. In *Interaction Design: Beyond Human-Computer Interaction* (pp. 37–67). John Wiley & Sons, Inc.

Shen, Y., Jiang, S., Chen, Y., & Campbell, K. D. (2022). *To Explain or Not to Explain: A Study on the Necessity of Explanations for Autonomous Vehicles* (arXiv:2006.11684). arXiv. http://arxiv.org/abs/2006.11684

Steinhauser, K., Leist, F., Maier, K., Michel, V., Pärsch, N., Rigley, P., Wurm, F., & Steinhauser, M. (2018). Effects of emotions on driving behavior. *Transportation Research Part F: Traffic Psychology and Behaviour*, *59*, 150–163. https://doi.org/10.1016/j.trf.2018.08.012

Stojmenova, K., Sanghavi, H., Nadri, C., Ko, S., Tomažič, S., & ... (2020). Use of spatial sound notifications for takeover requests in semi-autonomous vehicles-a cross-cultural study. In M. Zdravković, Z. Konjović, & M. Trajanović (Eds.), *ICIST 2020 Proceedings* (pp. 32–35).

Summala, H. (1981). Driver/Vehicle Steering Response Latencies. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *23*(6), 683–692. https://doi.org/10.1177/001872088102300605

Swain, R., Kaye, S. A., & Rakotonirainy, A. (2023). Is my AV crashing? An online photo-based experiment assessing whether shared intended pathway can help AV drivers anticipate silent failures. *Ergonomics*, *0*(0), 1–15. https://doi.org/10.1080/00140139.2023.2176551

Tabone, W., De Winter, J., Ackermann, C., Bärgman, J., Baumann, M., Deb, S., Emmenegger, C., Habibovic, A., Hagenzieker, M., Hancock, P. A., Happee, R., Krems, J., Lee, J. D., Martens, M., Merat, N., Norman, D., Sheridan, T. B., & Stanton, N. A. (2021). Vulnerable road users and the coming wave of automated vehicles: Expert perspectives. *Transportation Research Interdisciplinary Perspectives*, *9*, 100293. https://doi.org/10.1016/j.trip.2020.100293

Tan, X., & Zhang, Y. (2022). The effects of takeover request lead time on drivers' situation awareness for manually exiting from freeways: A web-based study on level 3 automated vehicles. *Accident Analysis & Prevention*, *168*, 106593. https://doi.org/10.1016/j.aap.2022.106593

Taylor, S., Wang, M., & Jeon, M. (2023). Reliable and transparent in-vehicle agents lead to higher behavioral trust in conditionally automated driving systems. *Frontiers in Psychology*, *14*, 1121622. https://doi.org/10.3389/fpsyg.2023.1121622

Telpaz, A., Rhindress, B., Zelman, I., & Tsimhoni, O. (2015). Haptic seat for automated driving: Preparing the driver to take control effectively. *Proceedings of the 7th International Conference on Automotive User Interfaces and Interactive Vehicular Applications*, 23–30. https://doi.org/10.1145/2799250.2799267

Tufte, E. R. (Ed.). (2013). *Envisioning information* (14. print). Graphics Press.

van de Merwe, K., Mallam, S., & Nazir, S. (2022). Agent Transparency, Situation Awareness, Mental Workload, and Operator Performance: A Systematic Literature Review. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 001872082210778. https://doi.org/10.1177/00187208221077804

Vogelpohl, T., Kühn, M., Hummel, T., Gehlert, T., & Vollrath, M. (2018). Transitioning to manual driving requires additional time after automation deactivation. *Transportation Research Part F: Traffic Psychology and Behaviour*, *55*, 464–482. https://doi.org/10.1016/j.trf.2018.03.019

Vogelpohl, T., Kühn, M., Hummel, T., & Vollrath, M. (2019). Asleep at the automated wheel—Sleepiness and fatigue during highly automated driving. *Accident Analysis & Prevention*, *126*, 70–84. https://doi.org/10.1016/j.aap.2018.03.013

Wang, M., Hock, P., Chan Lee, S., Baumann, M., & Jeon, M. (2022). Jarvis in the car: Report on characterizing and designing in-vehicle intelligent agents workshop. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *66*(1), 948–952. https://doi.org/10.1177/1071181322661445

Wang, M., Hock, P., Lee, S. C., Baumann, M., & Jeon, M. (2021). Genie vs. Jarvis: Characteristics and Design Considerations of In-Vehicle Intelligent Agents. *13th International Conference on Automotive User Interfaces and Interactive Vehicular Applications*, 197–199. https://doi.org/10.1145/3473682.3479720

Wang, M., Lee, S. C., Montavon, G., Qin, J., & Jeon, M. (2022). Conversational Voice Agents are Preferred and Lead to Better Driving Performance in Conditionally Automated Vehicles. *Proceedings of the 14th International Conference on Automotive User Interfaces and Interactive Vehicular Applications*, *1*(1), 86–95. https://doi.org/10.1145/3543174.3546830

Wang, M., Lee, S. C., Sanghavi, H. K., Eskew, M., Zhou, B., & Jeon, M. (2021). In-Vehicle Intelligent Agents in Fully Autonomous Driving: The Effects of Speech Style and Embodiment Together and Separately. *13th International Conference on Automotive User Interfaces and Interactive Vehicular Applications*, 247–254. https://doi.org/10.1145/3409118.3475142

Wang, M., Park, S. H., Lee, S. C., Hock, P., & Baumann, M. (2022). Build Your Own Genie and Jarvis: 2nd Workshop on Characteristics and Design Considerations of In-Vehicle Intelligent Agents. *14th International Conference on Automotive User Interfaces and Interactive Vehicular Applications*, 176–178. https://doi.org/10.1145/3544999.3552313

Waytz, A., Cacioppo, J., & Epley, N. (2010). Who sees human? The stability and importance of individual differences in anthropomorphism. *Perspectives on Psychological Science*, *5*(3), 219–232. https://doi.org/10.1177/1745691610369336

Waytz, A., Heafner, J., & Epley, N. (2014). The mind in the machine: Anthropomorphism increases trust in an autonomous vehicle. *Journal of Experimental Social Psychology*, *52*, 113–117. https://doi.org/10.1016/J.JESP.2014.01.005

Wiegand, G., Eiband, M., Haubelt, M., & Hussmann, H. (2020). "I'd like an Explanation for That!"Exploring Reactions to Unexpected Autonomous Driving. *22nd International Conference on Human-Computer Interaction with Mobile Devices and Services*, 1–11. https://doi.org/10.1145/3379503.3403554

Williams, K., & Breazeal, C. (2013). Reducing driver task load and promoting sociability through an Affective Intelligent Driving Agent (AIDA). *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, *8120 LNCS*(PART 4), 619–626. https://doi.org/10.1007/978-3-642-40498-6_53

Wong, P. N. Y., Brumby, D. P., Babu, H. V. R., & Kobayashi, K. (2019). "Watch out!" Semi-autonomous vehicles using assertive voices to grab distracted drivers' attention. *Conference on Human Factors in Computing Systems - Proceedings*, 5–10. https://doi.org/10.1145/3290607.3312838

Wright, J. L., Chen, J. Y., Barnes, M. J., & Hancock, P. A. (2017). *Agent Reasoning Transparency: The Influence of Information Level on Agent Reasoning Transparency: The Influence of Information Level on Automation-Induced Complacency*. *June*.

Xu, W. (2021). From automation to autonomy and autonomous vehicles: Challenges and opportunities for human-computer interaction. *Interactions*, *28*(1), 48–53. https://doi.org/10.1145/3434580

Young, K. L., Koppel, S., & Charlton, J. L. (2017). Toward best practice in Human Machine Interface design for older drivers: A review of current design guidelines. *Accident Analysis and Prevention*, *106*, 460–467. https://doi.org/10.1016/j.aap.2016.06.010

Zahabi, M., & Kaber, D. (2018). Effect of police mobile computer terminal interface design on officer driving distraction. *Applied Ergonomics*, *67*, 26–38. https://doi.org/10.1016/j.apergo.2017.09.006

Zang, J. (2023). *The Effects of System Transparency and Reliability on Drivers' Perception and Performance Towards Intelligent Agents in Level 3 Automated Vehicles*. Virginia Tech.

Zhang, J., Shu, Y., & Yu, H. (2021). Human-Machine Interaction for Autonomous Vehicles: A Review. In G. Meiselwitz (Ed.), *Social Computing and Social Media: Experience Design and Social Network Analysis* (Vol. 12774, pp. 190–201). Springer International Publishing. https://doi.org/10.1007/978-3-030-77626-8_13

Zhang, Q., Jessie Yang, X., & Robert, L. P. (2021). What and When to Explain? A Survey of the Impact of Explanation on Attitudes Toward Adopting Automated Vehicles. *IEEE Access*, *9*, 159533–159540. https://doi.org/10.1109/ACCESS.2021.3130489

Zhang, T., Yang, J., Liang, N., Pitts, B. J., Prakah-Asante, K. O., Curry, R., Duerstock, B. S., Wachs, J. P., & Yu, D. (2020). Physiological Measurements of Situation Awareness: A Systematic Review. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *3*. https://doi.org/10.1177/0018720820969071

Zhang, Y., Wu, C., & Wan, J. (2016). Mathematical Modeling of the Effects of Speech Warning Characteristics on Human Performance and Its Application in Transportation Cyberphysical Systems. *IEEE Transactions on Intelligent Transportation Systems*, *17*(11), 3062–3074. https://doi.org/10.1109/TITS.2016.2539975

Złotowski, J., Proudfoot, D., Yogeeswaran, K., & Bartneck, C. (2015). Anthropomorphism: Opportunities and Challenges in Human–Robot Interaction. *International Journal of Social Robotics*, *7*(3), 347–360. https://doi.org/10.1007/S12369-014-0267-6/FIGURES/1

APPENDICES

Appendix A. Dialogue for the ride-sharing use case.

Legend: 🧑 Xiaoming | 🕵 Xiaoming's Agent | 👱Kim | 🤖 Kim's Agent

[Xiaoming arrives at the airport, tired and wants to check in to the hotel]

🧑→ 🕵 : "To the hotel, wake me up when we're there."

🕵: "OK, we'll drive to the L7 Hotel."

[15min into the drive, Kim joins; Xiaoming wakes up due to the interruption and noise]

🧑 "What's going on?"

🕵 "Don't worry, a new passenger joined because you booked a shared ride. Are you up for a chat, or do you want to rest?"

🧑 "No!"

[Avatars are negotiating silently; zones in the car are created.]

[Kim joins the car]👱: "To my friend Lin."

🤖 [calmly]: "Ok, we'll drive to Lin. The other person wants to sleep; please do not disturb her. Should we switch to chatting or continue whispering."

👱 chats with 🤖: "Sure, let's chat..."

[The car arrives at Xiaoming's hotel]:

🕵: [vibrates seatbelt] ... [no reaction] … [Phone is vibrating, light is flashing]… [no reaction]

👤🤖 sync info about the language they speak:

👤 [in Korean]: "Can you help me wake him up?"

[👱 tries to wake 👦 up, softly punches 👦, and now 👦 wakes up]

👤: "We're there"

👦 [in Chinese]: "Sorry, I was too tired"

👤 [in Korean]: "She said I'm sorry she was too tired."

🤖: "We're there."

[Kim leaves the car]. "By, have a nice trip [in Korean]."

[🤵 translates to 👦]

[👦 arrives at the hotel, 👤 gives some direction advice]: "Hotel is the street down; it's on the 7th floor."

Appendix B. Contextual Inquiry Experimenter Script

Introduction

Hello, my name is Manhua Wang, and this is my assistant Yujiao. Thank you for participating in our study. Today, we'll be interviewing you to better understand how you interact with the in-vehicle advanced driver assistance systems while you're driving as you would typically do. Our goal is to learn from your experiences and identify any areas for improvement.

Throughout the session, you'll be video and audio recorded for us to analyze the data further. We'll record your vehicle's instrumental panel instead of your face. Here is the consent form for you to review and sign. This is the same as the one I sent you via email. Please feel free to ask me any questions you have.

[Hand the participant the consent form for signature.]

Before we start our trip today, I have several basic questions about your experiences with the advanced vehicle features and how you use them.

What driving assistant features does your vehicle have?

1. Can you walk us through them? What are these features about, and how do it work?

2. How often do you use them?

3. [For longitudinal and lateral controlling systems, ask follow-up questions] Typically, when will you use this feature? Could you walk us through the process of activating it?

   3.1. Do you need a minimum speed to activate them (ACC and LKA)?

   3.2. What about the traffic or weather condition?

4. When you purchase your car, did anyone show you how to use these features?

5. Do you have any concerns or hesitations about this feature?
   5.1. If so, what improvements would you like to have?

Transition

Thank you for answering my questions. Before we go on the road, I would like to clarify some of the rules regarding our interview. You would drive normally as you typically do on the road. When you are going to manipulate any feature, we ask you to verbally communicate your thought process regarding why you decide to do so, just like you did when you walked us through your vehicle features. During your driving, I might ask you some questions regarding your actions. If you think it is not a good time to answer the question, please let me know so we can save that for later. Do you have any questions?

Contextual Inquiry

[Pay attention to their actions when using the feature]

1. I noticed that you just use/operate something over there. Why did you decide to activate/deactivate that feature?

[Also pay attention to any frustration and emotional reactions]

I noticed that you did XYZ when the system notified you, what is that for?

[Some ice-breaking questions.]

1. What information would you like to see displayed while the vehicle is in ADAS mode?

2. Can you describe any specific scenarios where you might want to override the ADAS mode? [What about pedestrian heavy road]

3. In what scenarios you wouldn't use ADAS?

4. What your general attitudes toward automated vehicles?

Debriefing

Thank you for driving us around and answering our questions. To end your study, we have some further questions.

1. Have you ever had a time when your vehicle feature turned itself off, but you did not know the reason? Could you describe that situation?

2. What did you do to solve your confusion when it happened?

Thank you again and this concludes our interview today. Here is your compensation and could you sign this payment sheet for me.

Appendix C. Expert Interview Questions

1 Introduction--

1. Could you describe your current research focus?
2. How many years have you worked in the area you mentioned above? [If you have multiple distinct research focuses, describe your experiences separately]
3. Could you also describe the relevance of your current research with promoting explainability/human understanding of the automated vehicles [If the answer provided in Question 1 already covers the answers to this question, please skip].

   3.1. How many projects or years have you worked on this topic?

2 General Concerns regarding Explainable Automated Vehicles—

1. What is your understanding of the explainable interfaces in the context of automated vehicles?
2. What components do you think should be included in the explainable interfaces for automated vehicles?
3. What differences do you see in explainable interfaces for automated vehicles that distinguish them from other AI-powered decision-support systems?
4. How do you perceive the explainable interfaces in automated vehicles in terms of safety, trust, and user experience?

3 Discussion on Specific User Challenges

In this section, we'll present two user challenges we identified from our previous research activities.

Scenario 1: Jennifer is a research scientist in the field of transportation safety. During the holiday time, he traveled with his friends and rented a new car. While having the automated driving feature engaged on the highway, the driver seat and steering wheel suddenly started to vibrate. They took the closest exit and pulled over to figure out why the vibration happened. After 15 minutes, they still could not figure out the reason. But the vibration suddenly stopped. Jennifer continued her journey with her friends but was very worried about their vehicle condition. So, they ended up switching to another rental car, which caused a delay in the planned trip.

1. What is your understanding of the problems or causes of the user's situation from your expert perspective?

2. Based on your expertise and knowledge, what are some improvements of the vehicle systems that can be made to address their challenges under this situation?
    2.1. Whether and how do the adjustments you just mentioned change if the driver is not fully attentive, meaning that they are distracted for certain reasons (e.g., texting, drowsiness)?

Scenario 2: Jake is a software engineer who is a big fan of luxury vehicles, especially those with advanced driving automation features. He recently bought a vehicle with an advanced driver assistance system that allows hands-free. One day, he was driving on the freeway and activated the hands-free feature when driving in the left lane. Jake felt a little bit drowsy and gradually experienced some microsleep. The vehicle detected Jake's drowsiness and alerted him with a mild tone. Jake did not respond to the system alert despite the volume of the tone increasing gradually. The vehicle started to slow down on the highway with hazard lights but without changing to the right lane. Suddenly, Jake was awakened by an urgent honking and realized that he was traveling too slowly in the left lane. Jake took over the control and speeded up to avoid a potential crash.

1. What is your understanding of the problems or causes of the user's situation from your expert perspective?
2. Based on your expertise and knowledge, what are some improvements of the vehicle systems that can be made to address their challenges under this situation?

Follow-up questions after you complete the questions for two scenarios:

1. What are some other scenarios you could think of that can cause human confusion in using advanced driver assistance systems mentioned in the above two scenarios or higher levels of driving automation systems?
2. If you mentioned different approaches/methods to address user challenges described in Scenarios 1 and 2, what would be a better way to integrate them into a holistic, explainable system?

4 Trends in Explainable Automated Vehicles—

1. What are some risks you envision about XAI?

    1.1. How do you see a proper solution to address these pitfalls?

2. What are some essential design principles or frameworks to create effective explainable interfaces for automation systems? [If you also have opinions about automated vehicles, please include them.]

    2.1. Are there any specific methods or approaches that you find promising or handy in the context of explainable interfaces for automated vehicles?

3. What do you see as the future trends in the development of explainable interfaces for automation systems? [If you also have opinions about automated vehicles, please include them.]

4. Are there any areas of research that you believe require more attention or exploration for explainable interface automation systems? [If you also have opinions about automated vehicles, please include them.]

Appendix D. Events Description and Agent Explanations used in the Simulator Study

Scenario A

					Explanations		
No.	**Automation Mode**	**Environment**	**Event Description**	**Event Category**	**L2 + L3**	**L2**	**L3**
1	Auto	Highway	Lead vehicle with low speed		Slow moving car ahead, I will change lanes and pass it in a few seconds.		
2	Auto	Highway	Stopped car on the highway		Stopped car with flashing lights ahead, I will slow down and change lanes to avoid it in a few seconds.		
			Fog Ahead	Takeover Request	Heavy fog detected ahead, I won't work well due to limited camera visibility, please takeover when you're ready.		
3	Manual	Rural, Fog	Crash Ahead		Static obstacles in your lane ahead, your lane will be blocked in 600 feet.	Static obstacles in your lane ahead.	Your lane will be blocked in 600 feet.
4	Manual	Rural, Fog	Slow truck with hazard lights		Slow moving obstacles ahead, you might have to slow down in a few seconds.	Slow moving obstacles ahead.	You might have to slow down in a few seconds
			Clear Fog	Handover Request	Improved weather conditions detected, please reengage the auto-drive.		
5	Auto	Rural	Police car from opposite direction		Police car approaching with emergency, I will slow down to let it pass.		
6	Auto	Rural	Cyclists		Cyclists sharing the road ahead, I will slow down and change lanes to pass with caution.		
			Entering City	Takeover Request	Entering city ahead, I won't work well in complicated situations, please takeover		
7	Manual	City	Car pull over to the right		Stopped car with flashing lights ahead, your lane will be blocked in 500 feet.	Stopped car with flashing lights ahead.	Your lane will be blocked in 500 feet
8	Manual	City	Pedestrian Area		Busy pedestrian area ahead, you will encounter multiple crosswalks along this way.	Busy pedestrian area ahead.	You will encounter multiple crosswalks along this way

Scenario B

					Explanations		
No.	**Automation Mode**	**Environment**	**Event Description**	**Event Category**	**L2 + L3**	**L2**	**L3**
1	Auto	Rural	Dog crossing the road		Animals crossing ahead, I will reduce speed to avoid collision.		
2	Auto	Rural	Joggers on the road		Running pedestrians ahead, I will slow down and change lanes to pass with caution.		
			Entering City	Takeover Request	Entering city ahead, I won't work well in complicated situations, please takeover		
3	Manual	City	Red light		Traffic light changing ahead, our estimated waiting time is about one minute.	Traffic light changing ahead.	Our estimated waiting time is about one minute
4	Manual	City	Car running the red light		Car to the left is not stopping for the red light, you will be blocked in a few seconds.	Car to the left is not stopping for the red light	You will be blocked in a few seconds.
			Exiting City	Handover Request	Reduced traffic density detected, please reengage the auto-drive.		
5	Auto	Rural	Car traveling under low speed		Slow moving car ahead, I will change lanes and pass it in a few seconds.		
6	Auto	Rural	Construction		Entering the work zone, I will slow down and navigate through it shortly		
			Fog ahead	Takeover Request	Heavy fog detected ahead, I won't work well due to limited camera visibility, please takeover		
7	Manual	Highway, Fog	Crash on highway		Static obstacles in your lane ahead, your lane will be blocked in 600 feet.	Static obstacles in your lane ahead.	Your lane will be blocked in 600 feet.
8	Manual	Highway, Fog	Speed sign change		Traffic sign changing ahead, you might need to adjust for new driving conditions shortly.	Traffic sign changing ahead	You might need to adjust for new driving conditions shortly

Scenario C

					Explanations		
No.	**Automation Mode**	**Environment**	**Event Description**	**Event Category**	**L2 + L3**	**L2**	**L3**
1	Auto	Highway	Disabled car on highway		Disabled cars in your lane detected, I will slow down and change lanes to pass them in a few seconds		
2	Auto	Highway	Light rain		Light rain ahead detected, my system will adjust accordingly.		
			Fog ahead	Takeover Request	Heavy fog detected ahead, I won't work well due to limited camera visibility, please takeover		
3	Manual	Highway, Fog	Car traveling under low speed		Slow moving obstacles ahead, you might have to slow down in a few seconds.	Slow moving obstacles ahead.	You might have to slow down in a few seconds.
4	Manual	Highway, Fog	Police car from behind		Police cars approaching from behind, you should slow down or move over in a few seconds.	Police cars approaching from behind.	You should slow down or move over in a few seconds
				Handover Request	Improved weather conditions detected, please reengage the auto-drive.		
5	Auto	Rural	Road damage		Uneven surface detected ahead, I will change lanes to avoid vehicle damage.		
6	Auto	Rural	Dear walking aside		Wildlife near the roadway, I will slow down and pass them with caution.		
				Takeover Request	Entering city ahead, I won't work well in complicated situations, please takeover		
7	Manual	City	Car cutting into traffic		Car ahead attempting to merge back into the traffic flow, you might be cut off in a few seconds.	Car ahead attempting to merge back into the traffic flow.	You might be cut off in a few seconds.
8	Manual	City	Two-way stop intersection		Complete stop required ahead, the crossing traffic ahead will not stop.	Complete stop required ahead	The crossing traffic ahead will not stop

Appendix E. Semi-Structed Interview after Each Condition in the Simulator Study

1. What's your general perception of the intelligent agent you just experienced, and the information provided to you?

2. Which aspect are you satisfied with the information provided to you? Could you please elaborate?

3. Which aspect are you not satisfied with?

4. What additional information would you like to have when the car is driving?

5. What additional information would you like to have when you are driving?