CHALMERS UNIVERSITY OF TECHNOLOGY | UNIVERSITY OF GOTHENBURG

In-car voice assistants: The need and potential for AI-enabled voice assistants in vehicles

Master's thesis in Interaction Design & Technologies

AMANDA LIND JONSSON
ELIN HÖGMAN MÖLLER

Department of Computer Science and Engineering
CHALMERS UNIVERSITY OF TECHNOLOGY
UNIVERSITY OF GOTHENBURG
Gothenburg, Sweden 2024

Master's thesis 2024

In-car voice assistants: The need and potential for AI-enabled voice assistants in vehicles

AMANDA LIND JONSSON
ELIN HÖGMAN MÖLLER

UNIVERSITY OF GOTHENBURG

CHALMERS
UNIVERSITY OF TECHNOLOGY

Department of Computer Science and Engineering
Chalmers University of Technology
University of Gothenburg
Gothenburg, Sweden 2024

A Chalmers University of Technology Master's thesis template for $\LaTeX$
A Subtitle that can be Very Much Longer if Necessary
AMANDA LIND JONSSON
ELIN HÖGMAN MÖLLER

Supervisor: Ilaria Torre, Interaction Design and Software Engineering
Advisor: Jenny Wilkie, Volvo Cars
Examiner: Morten Fjeld, Interaction Design and Software Engineering

Master's Thesis 2024
Department of Computer Science and Engineering
Chalmers University of Technology and University of Gothenburg
SE-412 96 Gothenburg
Telephone +46 31 772 1000

Typeset in $\LaTeX$
Gothenburg, Sweden 2024

In-car voice assistants: The need and potential for AI-enabled voice assistants in vehicles
AMANDA LIND JONSSON
ELIN HÖGMAN MÖLLER
Department of Computer Science and Engineering
Chalmers University of Technology and University of Gothenburg

Abstract

This thesis investigates the role of AI-enabled voice assistants within the automotive industry, with a focus on user needs and regional differences between the United States and Sweden. Utilizing a mixed-method approach inspired by the TripTech method, this study collected both quantitative and qualitative data through surveys and expert interviews. The results revealed that more than half of the participants viewed in-car voice assistants (VAs) positively, valuing their safety and convenience for handling tasks and accessing information while driving. However, qualitative insights highlighted flaws in language comprehension and accuracy in current automated dialogue system, such as misdialed contacts due to misunderstood commands. While users from the United States generally perceived less importance in problem scenarios and were less likely to use VAs compared to Swedish users, the potential for generative AI to enhance VAs was recognized. This could lead to improved comprehension and more dynamic user interactions, although concerns about privacy and response accuracy persist. The findings suggest that while generative AI offers promising enhancements for in-car VAs, significant challenges in privacy and accuracy need to be addressed to fully leverage this technology in vehicles.

Keywords: Voice Assistant, Generative AI, Large Language Models, Automotive, Triptech method, user needs, survey, interaction design.

Acknowledgements

We would like to express our sincerest appreciation to our case company Volvo Cars for giving us this exciting thesis opportunity as well as to all people at Volvo Cars who have supported us throughout the project. We would also like to express a special thanks to our supervisor Jenny Wilkie for providing us with invaluable feedback and support. Lastly, immense thanks to our academic supervisor Ilaria Torre for her continuous guidance and feedback throughout the project.

Aamnda Lind Jonsson & Elin Högman Möller, Gothenburg, 2024-06-05

Contents

List of Figures

List of Tables

1

Introduction

Voice User Interfaces (VUIs) are quickly becoming more popular in the consumer space, embodied by digital assistants (eg. Amazon Alexa, Google Home) as well as Virtual Assistants found in mobile devices and cars [1]. The growth in popularity has prompted various industries such as the automotive industry, travel agencies, financial service providers and appliance manufacturers to integrate these interfaces into their products and services to meet customer expectations. This integration enables seamless interaction with customers through voice, facilitating user access to services and appliance control through voice commands. Dialogue systems in the form of digital assistants, are available to users across devices and in different contexts from smart speakers in the home to mobile devices on the go. These assistants can provide information in several topics such as about the weather, maps, schedule, call, events, etc [2].

The underlying technology supporting these dialogue systems are rapidly developing. The emergence of Generative Artificial Intelligence (AI), particularly Large Language Models (LLMs) have instigated a paradigm shift [3]. With their expanding capabilities, LLMs have introduced new user interfaces, conversational-based service systems, and quickly attracting a wide audience. Services like ChatGPT and Gemini, known for their advanced understanding and communication skills, deep world knowledge, and reasoning abilities, are well-prepared to help users with a range of daily and professional tasks.

This shift has introduced new technology providers into the Voice Assistant (VA) ecosystem which presents the era of next-generation conversational systems offering advanced linguistic capabilities and personalized multi-turn interactions [3]. While adopting these new language technologies does present its challenges, it also offers companies an opportunity to invest in, explore and integrate new applications with engaging User Experiences (UX). Moreover, the value of collecting and extracting insights from natural language data derived from conversations with users cannot be overstated, as it contributes significantly to enhancing product and service offerings. Amidst the technical advancement in the field lies a critical gap in understanding the alignment between user expectations and the practical utility of these interfaces. A broader understanding is needed about how new natural language technology enablers and emerging assistant ecosystems can be leveraged to create value for both the customer and the business.

1.1 Purpose & research question

The purpose of this master thesis is to explore innovative interaction models for AI-enabled conversations with users in automotive contexts. It aims to understand user needs, challenges and opportunities of these technologies in creating engaging customer-facing applications and improving UX. Further, the study aims to explore the attitudinal differences towards using in-car voice assistants between individuals in the United States and Sweden, investigating if different driving cultures have an impact on user needs and preferences. The research will investigate the issues drivers face, and propose solutions through voice user interfaces. Additionally, the thesis will assess how the design and functionality of these interfaces impact their utility and user satisfaction, ultimately aiming to enhance the automotive UX. The primary goal is to bridge the gap in understanding how these advanced technologies meet user expectations and offer practical utility in real-world applications. The thesis will also explore how to effectively leverage these emerging technologies to create value for both customers and businesses in the automotive industry. The thesis contributes to the field of interaction design by providing insights into innovative AI-enabled interaction models for in-car voice user interfaces, assessing their impact on user experience, and exploring cross-cultural attitudinal differences, thereby offering valuable guidance for the development of user-centric automotive technologies.

The research questions are proposed as

- Is there a need for AI-enabled in-car voice user interfaces, and what specific functionalities are drivers seeking?
  - Are there any differences in attitudes towards using in-car voice assistants between individuals in the United States and Sweden?
- Does Generative AI have any potential in improving in-car Voice Assistants?

2

Background

This chapter explores the evolution of voice user interfaces and highlights current trends in In-car voice assistant technology.

2.1 Evolution of Voice User Interfaces

The history of in-car Voice User Interfaces (VUIs) has seen continuous advancements to enhance driver safety and convenience [4]. In the 1990s and early 2000s, initial efforts focused on simple command and control actions with limited vocabularies, designed to work under various noise conditions. Early projects like SPEECHDAT-CAR in 1998 aimed to gather diverse speech data in different in-vehicle conditions to support the development of automatic speech recognition (ASR) systems [5].

In the early 2000s, projects like CU-Move made further progress by developing interactive in-vehicle speech systems [6]. These projects analyzed in-car noise conditions and collected speech data, highlighting the impact of factors like open windows and varying speeds on speech recognition accuracy. The AT&T-DARPA Communicator Project in 2000 created robust multimodal dialogue systems, setting the stage for intelligent in-vehicle systems [7].

Commercialization began with systems like Ford Sync (2007-2012), which made voice recognition a mainstream feature, allowing drivers to control various functions through voice commands [4]. The 2010s saw the rise of cloud-based intelligent personal assistants like Apple Siri, Google Assistant, and Amazon Echo, which influenced in-vehicle speech systems by leveraging cloud computing for high language-understanding accuracy and a wide range of applications, though they required reliable internet connectivity.

Despite the advancements, there is still room for improvement in VUIs [4]. Future systems aim to provide more natural interactions and better integration with other in-vehicle systems. The evolution of in-vehicle speech recognition and dialogue systems shows significant interest in incorporating advanced AI and machine learning technologies to improve user experience and safety in smart vehicles.

2.2 Trends in In-car voice assistant technology

In a report from Capgemini Research Institute [8], the attention is drawn to how automobile manufacturers can deliver an enhanced in-car voice experience. The research highlights the increasing relevance of voice and chat assistants in engaging customers and organizations. These in-car assistants were initially tasked with basic functions, but are now rapidly evolving in intelligence. This progress has caught the attention of car manufacturers, who are now dedicating a lot of resources to enriching the UX with these VAs. The study presented an excitement for the implementation of the technology in cars. However, it also emphasized a need for enhancements in the overall experience provided by these in-car VAs. This feedback has generated interest within the automotive domain to ensure that voice services are not only reliable but also relevant. Implementing such improvements would enable car manufacturers to align with user expectations and thereby foster increased brand loyalty. Moreover, a research conducted by YouGov [9] in Great Britain, found that over a third (35%) of the British people have already used a VA in a car before, with 9% saying they do so often. Additionally, 38% said they are interested in having a VA in their next vehicle, indicating a growing interest of utilizing in-car VA technologies.

Over the recent years, the technology has evolved rapidly; in the beginning of January 2024 the Customer Electronics Show (CES), recognized as one of the largest influential technology events in the world, presented innovative solutions aimed at improving in-car UX through VAs developed by various car brands [10]. Among many notable innovations released at the CES event, a selection of advancements were focused on improving in-car UX. Some of the noteworthy examples include Volkswagen, Mercedes-Benz and BMW. Volkswagen unveiled a vehicle equipped with ChatGPT, an AI-based chatbot integrated into their VA. The technology was integrated to facilitate seamless, hands-free interactions [11]. It is said that this feature enhances conversations, provides answers to questions, and delivers vehicle-specific information, among other capabilities, all this through intuitive language. Furthermore, Mercedes-Benz [12] introduced its latest technologies. They presented a new Mercedes-Benz User Experience (MBUX) virtual assistant, leveraging generative AI and 3D graphics into the vehicle to foster user-vehicle interaction. The assistant is aimed to offer natural, intuitive, and personalized interactions, allowing the user to communicate with the system easily and effectively. Moreover, BMW demonstrated a development demo of their Intelligent Personal Assistant, enhanced with generative AI and Amazon Alexa Large Language Model [13]. BMWs first Intelligent Personal Assistant was launched in 2018, however, this updated assistant offers rapid, natural interactions, detailed information on vehicle functionality, and some control of the car, aiming to improve safety and enhance the driving experience.

To summarize, the integration of more versatile and advanced VAs in cars has been an ongoing request due to the shortcomings of current VAs in meeting user expectations. Nowadays, technologies related to VAs and their capabilities are improving quickly, moving closer to meet the users expectations and needs, making the driving experience more enjoyable.

3

Theory

This chapter introduces the theoretical part of the study. It aims to clarify and explain the terminology and concepts used throughout the study. The section starts with discussing human agent interactions, followed by voice user interface, human vehicle interactions and conversational design. The chapter also covers Generative AI as well as the technologies behind large language models and text-to-speech systems. Lastly, it discusses challenges in this area.

3.1 Human Agent Interaction

Human Agent Interaction (HAI) refers to some kind of direct communication between people and embodied AI. The idea of AI has many different interpretations [14]. However, the AI pioneer Nils Nilsson suggests defining intelligence as a quality that allows an entity (or agent) to act appropriately and with awareness of its surroundings [15]. This entity, whether artificial or natural, isn't restricted to specific skills, tasks, technologies, or physical appearances according to this definition. Embodied AI refers to AI taking a tangible form, such as a robot or agent that allows for physical or sensory interaction [16]. However, the concept of embodiment extends beyond just physical presence to also considering how closely these forms resemble or differ from living beings in their design and function. The varying degrees of embodiment, from basic physical forms to fully organic systems, showcase the range of AI's potential interactions with the real world.

Embodied AI generally refers to a physical body of which the user can interact, Conversational Agents (CAs) can be both embodied and disembodied or virtual [17]. The scope of this study mainly focuses on virtual CAs. CAs are dialogue systems with a great variety of possible applications. In its most basic form, the dialogue systems are meant to recognize speech or text, handle the interaction and give information to the user [18]. The CAs have integrated into everyday life, acting as digital intermediaries that transform the relationship between businesses and customers where they serve as a bridge between them. These CAs include text-based conversational agents such as OpenAIs ChatGPT and Google's Gemini as well as speech based conversational agents such as Google Assistant, Alexa and Siri [17], [19]. These conversational agents are known as disembodied or virtual as they are not situated in a physical body where the user can interact. Rather the agent is situated in the software of another product's hardware, such as a speaker or computer [18]. With the current developments in research on AI algorithms, Natural Language Processing (NLP) and LLMs, the capabilities of the CAs have expanded. These AI-powered agents can comprehend and generate natural language, learn from experiences, and simulate emotions.

3.2 Voice User Interface

Voice is a medium of communication used to carry information via sound from a speaker to one or many receivers. We commonly think of speech when thinking of voice, a spoken language articulated by what is linguistically known as a combination of vowels and consonant sounds [14]. Speech, however, is only one part of the voice. Voice also uses non-linguistic attributes to send signals about the speaker's emotional state, personality and gender. Voice design is studied within several domains such as human-robot interaction, psychology, linguistics and phonetics [20]. Current research in these domains is focusing on the topic of voice design, particularly examining auditory cues like pitch, accent, and speech rate to understand how they impact on the behavior and perception of receivers. With the emergence of new technology such as AI, voice is now being used as a medium to communicate with vehicles, computers or smart speakers. In recent years, many new Virtual Assistants and smart speakers have increased in popularity [14]. These products make use of the voice to simplify the interaction with the device and allow the user to complete complex tasks in a simplistic way. VUI is defined as a user interface that takes speech as an input using speech recognition and gives a spoken output generated by technologies like speech synthesis [21]. In other words, it is a way for people to talk to devices using their voice. The device listens to what you say and can respond with spoken words or recorded messages. This kind of interaction can be very beneficial when the users hands or eyes are occupied by other activities, such as when driving a car.

3.2.1 Learnability VUI

When technologies such as VUIs emerge, it is important to consider how much the user needs to know beforehand and how much they need to learn to use these technologies [22]. Users often encounter usability problems with VUIs, such as poor feedback and errors in the system, which can cause users to struggle when using it. Users may not know if the system failed to understand them or simply did not hear them correctly, which can negatively impact the overall UX. However, over time and practice, users can develop better understanding of how these systems work and what to expect, this in turn can reduce some of the initial issues. Despite this, it is crucial for designers to consciously improve these systems. By addressing these usability problems and enhancing the systems ability in more human-like and reliable conversations, these designers can significantly reduce the learning curve for users.

3.3 Human vehicle interaction

Human Vehicle Interaction (HVI) is strongly correlated to the field of HAI. It aims at understanding and enhancing how people and vehicles interact and communicate [23]. The domain of interaction specifically studies how humans sense, perceive, share information, make inferences and decisions while interacting with the vehicle. As the vehicles become more automated, the drivers will have more time and possibility to perform other tasks, leading to new ways of interacting with the vehicle. This development has led to many automotive manufacturers including VAs in their vehicles [24]. These digital agents offer benefits for both drivers and vehicle manufacturers. For drivers, the agent can support the user in their driving performance and safety as well as improving the driving experience [24], [25]. They can support drivers performing tasks and accessing information while driving. They also meet functional needs, such as responding to commands, and can be used as a social conversation partner improving the driver's emotional state.

Driving demands a lot of focus on the surroundings. Using a VA can reduce distractions and improve safety in traditional driving as it lowers the cognitive load by handling non-driving tasks such as handling phone calls and music media entertainment [26]. For car manufacturers the main benefit is the relationship built between the VA and the customer [18], [24]. This relationship can promote customer purchase intentions, increase car sales and improve customer satisfaction. Additionally, the data collected from the VAs contain valuable information about driver preferences and habits, and thereby understanding driver behavior.

3.4 Conversational design

Conversation is a dialogue between two or more people, involving generation and exchange of information. Conversations are essential for learning, coordination, and collaboration [27]. Effective conversations foster mutual understanding and consensus among the participants, which are fundamental in creating trust and building relationships. Conversational design is the dialogue between a human and a digitally generated system, it is focusing on creating natural human conversations that consider both human needs and technological capabilities, aiming to foster beneficial and dynamic interactions.

In designing for conversational design it is important to consider several aspects. It is critical to identify the user and the use cases for the conversation. This can be achieved by developing a system persona, different types of user personas, and understanding the context and desired goals of system use. This, along with the insights of the technological capabilities and limitations lays the foundation of how designers should proceed in creating valuable dialogues [27], [28]. Once the clear vision of who is using the system and in which context it will be used is established, sample dialogues could be designed to explore new features and strategies of how to shape the conversation efficiently [28].

To ensure natural interactions between the system and the user, the system must recognize the context of the information that the user is providing, as well as managing follow-up conversations [28]. Further, creating a system's variation of response phrases can prevent a robotic feeling and instead contribute to natural conversations. Implementing turn-taking cues is an approach to inform the user when it is their turn to speak, maintaining the conversation in sync without interruptions. It is assumed that users are generally informative and cooperative, the system should therefore respond accordingly. However, there is no guarantee that the system will be able to handle cooperative responses effectively. Any arising errors should be addressed in a manner that lightly acknowledges the issue without fixating on it, focusing on maintaining the flow of the dialogue and redirecting the conversation. The goal is to acknowledge the mistake briefly and move on, rather than interrupt the entire conversation. Lastly, all system responses should remain relevant and without redundant information.

3.5 Generative Artificial Intelligence

Generative AI is a subset of AI technologies, utilizing machine learning, deep learning and other techniques to generate new content by analyzing data, information and patterns from the training data [29]. Figure 3.1 shows Generative AI and other AI-concepts [30]. In contrast to discriminative models which are designed to recognize and classify data, generative models learn the training data structure to generate new and unique instances of data that resemble the training data [29]. This capability makes Generative AI very powerful in a wide range of application areas, especially those including content creation and data augmentation. In November 2022, OpenAI introduced one of today's most renowned GenAI services, ChatGPT. ChatGPT is a chatbot leveraging advanced AI language technologies such as LLMs. Due to ChatGPTs ability to hold human-like conversations, it quickly increased in popularity and is now the fastest growing innovation, reaching 100 million active users within just two months.

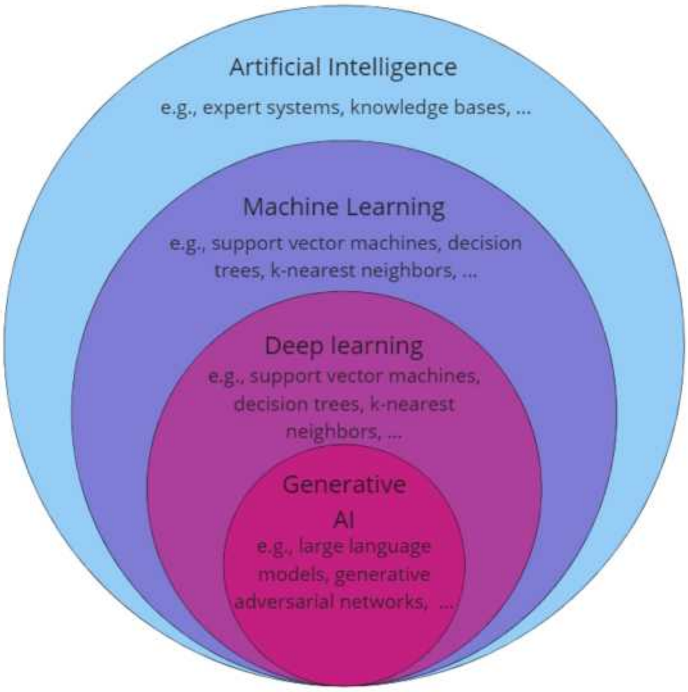


Figure 3.1: Generative AI and other AI concepts, inspired by Banh [30]

3.6 Large Language Models

LLMs are deep learning models and a type of AI that utilizes and is trained on massive datasets. They have the purpose of achieving levels of performance in understanding and generating general language tasks. An example of such a conversational-based service system is ChatGPT [31].

LLMs contain hundreds of billions of parameters and are built with a transformer-based architecture, which means that they manage to handle sequential data efficiently [31]. This deep neural network, also called transformer, has marked a significant breakthrough in language models due to its technology that has made LLMs better than other existing methods, highlighting their high-speed generation and improvement in the text quality they produce. A feature key of transformers is the self-attention mechanism which means that the models can weigh the importance of tokens, such as a word, parts of a word, or characters [31], [32]. This is done regardless of the tokens position in a sentence. The LLM can with this knowledge pay attention to how often a token appears in a specific order. It understands the statistical distribution of tokens and uses it to determine how likely one token is to follow another in a sequence. This feature can improve contextual understanding, generate more coherence and relevant text, produce more human natural responses with less fine-tuning, and increase the efficiency of generated language.

However, it has been demonstrated that while LLMs deliver excellent performance, they also come with certain limitations and challenges [32]. LLMs do not possess beliefs in the manner that humans do, which means that LLMs can not distinguish the truthfulness of the answers they provide. Their operation is based only on sequence prediction, identifying which tokens usually follow others based on their training data.

How the LLMs work could therefore lead to a result of leveraging heuristics or biases, where the model consistency focuses on a specific token within an input sequence, without considering the surrounding context [33]. This underscores the models reliance on learned patterns [34]. Such models inherently learn and replicate the biases present in their training data, could potentially perpetuate stereotypes or biases favoring certain groups or perspectives in the outcomes produced by the models.

Additionally, this may cause generating of fake or misleading content raising ethical concerns, or lead to hallucinations, where the model confidently produces high-quality text with incorrect information, convincing the user that the provided information is accurate [31]. However, given the models reliance on its training data, it cannot guarantee the accuracy or relevance of the information provided.

To address the issue of inaccurate information and improve the reliability of model outputs, methods have been developed to enhance the quality of these outputs [35]. A technique that has been used for LLMs is the Retrieval-Augmented Generation (RAG). It enables the model to access external sources beyond its initial training material, this could include specialized domains or other internal sources within organizational content. This approach is cost-effective for upgrading chatbot functionalities without the need of retrain foundation models. It also allows the Generative AI to be more accessible and practical and ensures the models to pull the latest research which improves their relevance and accuracy in their responses. Additionally, RAG can provide the user with the source material, enabling them to verify the information themselves. This flexibility and reliability make RAG a valuable tool for deploying AI across a wide range of applications with increased confidence.

3.7 Natural language generation

Traditional dialogue systems make use of decisions trees to determine how to respond to users [36]. This method involves manually creating a set of predefined rules and responses for different user input. This type of system can only respond to the exact questions that it has been programmed to respond to. For example, a user may ask what the weather is like today, then the dialogue system would follow a predetermined path to provide the weather information.

With the advancements in Natural Language Generation (NLG), a subfield within generative AI, new ways of constructing dialogue systems are possible. NLG is focused on generating text and in both written and spoken situations from structured data [37]. The NLG utilize AI models, such as LLMs, to generate responses dynamically based on the user input [38]. This technology can have a positive impact on businesses. NLG generates personalized text, customized to the individual user, which facilitates the interaction with the user through, for example, VAs and conversational chatbots. By providing users with personalized, relevant content and requested information, NLG can engage users, resolve issues quickly, and thereby build trust between the user and the business. These advantages, combined with precise and motivating language can enhance products and benefit businesses.

3.8 Automatic speech recognition

Automatic Speech Recognition (ASR) is a technology that enables computers to identify and process spoken language [39]. The ASR is trained to recognize the sound of human speech and convert it into text, preferably in the script of the spoken language, by using AI and machine learning algorithms. An ideal ASR system should be able to interpret the provided input and recognize the spoken words, and then utilize these recognized words as input for another machine to enable execution of some action based on this input [40]. This way of communicating with machines is predicted to be the future of human-machine communication.

3.9 Text-to-speech

Text-to-speech (TTS) is a technology that transforms text into spoken words. This involves training computers to understand and reproduce both written and spoken forms of language through the use of computer models [41]. This approach mimics the human process of learning and processing language. A key factor in improving TTS systems includes training with large amounts of data, such as LLMs [42]. By utilizing a comprehensive collection of diverse speech data, progress has been made in generating speech that closely resembles natural human speech. TTS technology operates in two main phases, the analysis of the text and the generation of speech waveforms. The incorporation of AI and machine learning has significantly advanced TTS, enhancing its ability to produce speech that closely resembles human vocalization.

Several key factors influence the effectiveness of TTS in communicating its message to users [41]. These include the voices appeal or preference, which affects how the users respond to the voice, intelligibility and clarity, referring to the ease with which words are understood and distinguished; naturalness, which measures how closely the speech resembles human speech, including emotional expressions and natural flow; and comprehensibility, indicating how easy it is for the users to follow the spoken words. Together, these aspects determine the overall UX with TTS technology.

3.10 Challenges

One of the challenges in providing a good UX of an in-car voice agent is determined by how good the network connectivity is. A study conducted by Wang [3] found that the UX is generally good when the network connectivity is good. Therefore, when using the VAs under poor network conditions the UX generally decreases drastically.

However, many of the current VAs use a hybrid storing technology, having some competence stored locally and some stored in the cloud [43]. This means that an internet connection may not be necessary to access all functions. For example, an offline VA could could handle car-related questions, such as providing guidance on how to connect to Bluetooth, while more complex tasks like discussing Harry Potter movies may require an internet connection [44]. With VAs partially stored locally, drivers will not have to worry about tunnels or dead zones affecting their interaction with the VA.

There are also risks concerning ethics, user privacy, and safety connected to the voice agents. As the VAs collect sensitive information about the driver, such as recordings from the microphone or location data, it is crucial to ensure that this information is secure [3]. Failure to secure this data can lead to privacy breaches, misuse of data, information leaks and tracking of individual behavior. Subsequently, this information can be sold for advertisement or commercial purposes.

Furthermore, algorithms used in speech or language models may be biased, potentially leading to discrimination of certain groups [3]. To prevent these biased LLMs researchers are exploring adversarial training of these models. Adversarial training refers to in its training phase, intentionally creating examples that are designed to challenge or mislead the model [45]. This example training data works alongside the original training data with the objective of making it more resilient to manipulative inputs and mitigating bias.

4 Methodology

This chapter introduces the chosen methodologies. It presents the theoretical framework Triptech utilized in the study. A mixed-methods approach was used to collect data, combining quantitative and qualitative methods, through surveys and interviews.

4.1 Case study

A case study is a detailed, in-depth research approach employed to a particular group or community, such as an organization. It aims to gather a deeper understanding of a specific problem [46], [47]. A case study can be structured by collecting qualitative or quantitative data, the strategy of this study employed a mix-methodology approach, incorporating both qualitative or quantitative data. This strategy involves analysis of both numerical and non-numerical data to gain insights of the problem, exploring the underlying concepts and experiences to generate new research ideas. While case studies offer significant advantages, they may face disadvantages and limitations. Case studies often generate large amounts of data, which can be difficult to analyze and organize, this often requires a well-planned strategy to optimize the work and the outcome. Aside from these challenges, case studies may encounter skepticism in the terms of credibility and generalization [48]. Bell & Waters points out the difficulties in verifying the collected data within case studies, which may result in misleading outcomes and raise concerns about the overall reliability and credibility of the research findings.

4.2 Theoretical framework: Triptech method

Nowadays, product innovations are often sprung from technical breakthroughs or advancements rather than from a customer need, making product development more technology-driven [49]. The products are developed upon the assumption that there is a user need for a certain technology integrated into a product. A common way to test whether these user needs assumptions are accurate is to measure the UX. Triptech is a UX research method for evaluating early design concepts. This method can be used both when evaluating design concepts in their initial stage as well as tech-first innovation practices. Triptech is a combination of three central perspectives, triangulation, prototype and technology [50]. Triangulation refers to using many research methods to validate user needs. These methods involve interviews, surveys, and group discussions, to collect data and get insights concerning user requirements and behaviors. Prototyping includes rapid prototyping techniques to test new designs and ideas as well as validating assumptions about user behaviors or needs. The technology part of Triptech involves using the appropriate software tools, collaboration platforms and management applications to streamline the process of design and development. The Triptech method is carried out in two main parts. The first one being surveys and the second one is focus groups. These two parts will be explained in more detail below.

The surveys are used to evaluate the user needs. Ideas developed from a technological viewpoint are usually based on hypotheses about how they might meet specific user needs [51]. The surveys force the team to create a clear and concise articulation of how the need or problem is meant to be addressed. 'What is it that we think we are able to address with this new technology?'. The surveys also provide results that can be generalized, used as an initial filter to identify and prioritize user needs and concepts [49]. They aim to address the frequency and importance of the identified problem or need. This method is particularly well suited when there are more problems or needs than what can be tested with only one person. The scenarios in the survey are generally constructed as the following example 'When I hear music I like, I want to know the artist and title (e.g. a song playing in a coffee shop)'. Structuring the scenarios like this puts the focus on a single user need problem. Figure 4.1 presents how the Triptech method measures the desirability of a design concept, function or feature.

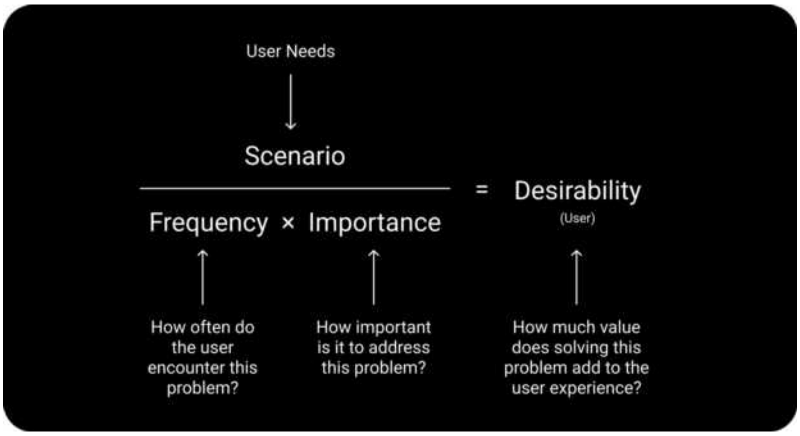


Figure 4.1: How to assess user needs in the early stages of product development by [50]

The purpose of the group interview is to evaluate the proposed solution to the problem or need [49]. The focus group methodology is carried out by combining three different methods, storyboarding, focus groups and speed dating. Speed Dating (SD) is continuously used throughout this second part of the Triptech method. The purpose of using SD is to quickly explore a wide range of ideas and concepts with participants to gather insights on their needs, preferences, and reactions. Showing design concepts to participants where they can engage in brief discussions and share thoughts.

The storyboards presented in the focus groups contain three frames, the first one shows the same problem statement as used in the surveys [51]. The second frame shows the solution or design concept intended to address the problem. The third frame presents the outcome of the solution, emphasizing the value it delivers to the user. By presenting the storyboards, participants can give feedback and evaluate the need and the design concepts. This is done in two steps; firstly individual feedback, and then focus groups. The individual feedback is important to get the participants first reactions to the needs and design concept, without influence from other participants. At this initial stage the participants are asked to rate the utility and desirability. After collecting the individual feedback, a group discussion is initiated within the focus groups. The purpose of this is to find out what the participants liked and disliked about the idea, get insights into possible use cases, as well as understanding what questions and concerns arise when seeing the design concept.

A notable limitation to this method is its reliance on participants to self-report their needs, rate the frequency and importance of their needs [49], [51]. An alternative to this approach could be, direct observation. Another limitation is that it uses focus groups, inheriting all the common limitations of using focus groups, such as group thinking and participants changing their answers after listening to other participants in the group.

4.3 Literature review

Conducting a literature review requires an effective search strategy to gather the most relevant and suitable sources [48]. A practical method involves clearly defining the scope of the literature review, where the limitations of the study are often determined by the research question that needs to be answered. Defining a scope is crucial for identifying and utilizing specific keywords to filter the search outcomes and to guarantee relevant sources findings relevant to the topics scope. It is also a good practice to utilize multiple sources to find relevant literature for the study.

4.4 Scenarios

A scenario is a description of potential future actions or events [49]. It can for example be used to gather better understanding and insights in how customers interact with a system or how a system could interact with the customer. Scenarios aim to imitate real-life situations in which users engage with the system. The purpose of examining these scenarios is to understand the users perspective and how the system could be used in diverse contexts, identify common patterns, challenges and areas of improvement. This approach aims to obtain a richer insight in the user's needs and expectations.

4.5 Survey

Surveys are often conducted to collect larger amounts of data [52]. It is an efficient and flexible data collection method that can be used for several types of research. Its main purpose is to capture needs, preferences and expectations to understand the general characteristics of a group of people. Furthermore, surveys can be designed to collect both qualitative data, resulting in numerical insights for statistical analysis, and quantitative data, that provides detailed free text responses from open-ended questions.

4.5.1 Participants

The participants in the surveys were all Volvo employees who are a part of the Co Development (CoDev) program at Volvo Cars. The CoDev program is a company car program that enable employees to actively take part in and support product enhancement and development. The primary aim of this program is to monitor, evaluate and gain insights through objective data collection, such as data collected from test equipment in vehicles, as well as subjective data provided by the drivers through questionnaires, forms, interviews or user studies. When employees become a part of the CoDev program they are required to sign a contract giving consent to data collection, in accordance with GDPR. The data will only be shared with employees, contractors or agents at Volvo Cars with a specific need for the information. The privacy notice, to which the driver has given consent upon joining the program, is linked on the first page of the survey as a reminder for those who may need it. The survey does not contain a section for giving consent for the collection of their data, as this has already been established when joining the program.

The initial plan was to focus on CoDev drivers from Sweden. However, during a meeting with the organizer of the survey distribution, it was mentioned that Co-Drev drivers are located in multiple countries, including the United States. This prompted us to expand our study and compare the interest of using VAs between the two countries. The United States was particularly chosen based on our personal hypothesis that there is a difference in the driving habits in the US compared to Sweden, such as more frequently waiting in traffic, which might affect the interest of an in-car VA. Additionally, statistics show that nearly half of the US population use VAs regularly, and the number is rising [53], indicating a high familiarity with this technology. Therefore, investigating any differences in attitude towards voice assistants became interesting.

4.6 Expert interviews

Expert interviews are conducted to collect relevant and unique insights from an expert in a specific field [54]. An expert is characterized by possessing specialized knowledge and they typically holds significant roles, often with unique access to information. The purpose of these interviews is to dig deeper into experts' insights and perspectives, using their special knowledge to develop a richer comprehension of the problem and explore solutions. Expert interviews are especially valuable in research to obtain nuanced insights that might not emerge from standard data collections. Through their expertise and understanding, experts can provide context and clarity when addressing complex issues.

4.7 Quantitative methods

The quantitative research methods are statistical techniques commonly utilized when trying to decode and interpret data that can be quantified, which means that the data is organized in a fixed and orderly manner [55], [56]. This type of research is central for drawing objective conclusions based on data that can be precisely measured such as percentages, frequencies, averages and correlations.

4.7.1 Descriptive statistics

In this study, two basic quantitative analysis techniques were used. These are the average and the percentages. The percentage is used to simplify the data, which facilitates comprehension and communication of the raw data [55]. Additionally, it standardizes the data which allows for comparison of different groups and categories. The average consists of three types, the mean, median and mode. The survey consisted of ordinal data which led to the use of the numerical measurements median and mode. Typically, the mean is more suited for continuous numerical data and was therefore not included in this analysis. The mode is used to represent the most commonly occurring value. The median is used to identify the central point of the ranked data, offering another perspective on the datasets central tendency.

4.7.2 Spearmans rank correlation

The Spearmans rank correlation is a non-parametric version of the Pearson correlation, used to measure the strength and direction of the relationships between two ranked variables. This method is useful when working with ordinal data, where the order of values is more important than the exact values. Spearmans correlation uses a monotonic function, which looks at two different relationships:

- If the value of one variable increases, the other value of another variable also increases
- If the value of one variable increases, the other value of another variable decreases

This method captures the directional association between two variables [57]. The result of the correlation goes between -1 and 1. Where -1 indicates a perfect negative correlation, 0 indicates no correlation and 1 indicates a perfect positive correlation. If there are tied ranks in the dataset (two or more identical values), each tied rank is given the same average rank.

Additionally, to conclude if the correlation is statistically significant (if the correlation is likely to have a true underlying effect or due to random variation), a p-value is calculated [58]. This study looks at the significance level of 0,05 and 0,01. Where anything equal to or below these numbers indicates a rejection of the null hypothesis. This result indicates that the observed correlation is strong enough and that there is only 5% or 1% (0,01 being a stricter criterion) possibility that the correlation is due to random chance.

4.7.3 Mann-Whitney U test

Mann-Whitney U-test is a non-parametric test to conclude if there are any differences between two independent groups [58]. It is a version of the original Student t-test but is usually used for ordinal data which is the case in this study. Furthermore, Mann-Whitney U test is robust against differences in sample sizes, making it suitable in this study. To determine if there is a statistical significance between the two chosen groups, the result should be equal to or lower to the significant level which typically is the p-value <= 0,05. The null hypothesis, which states that two independent groups are equal, is rejected at the significant level, meaning that one of the groups has lower or higher values than the other.

4.8 Qualitative methods

The qualitative research methods are used primarily to understand complex matters by analyzing non-numerical data [55]. This type of analysis is used when researchers seek rich and detailed insights into human behaviors, opinions and social processes. Qualitative analysis is typically favorable when the objective is to explore new topics where existing theory or research may be limited or when a deep, contextual understanding is required. Methods which are commonly used are data categorization, thematic analysis and critical incident analysis.

4.8.1 Thematic analysis

Thematic Analysis (TA) is a method for identifying, organizing and interpreting patterns of meaning (themes) in a dataset [59]. By focusing on the meaning across the entire dataset, TA allows researchers to find and understand shared views and experiences. The primary focus of TA is not to identify meanings and experiences that are unique to a single data item. Rather, it focuses on uncovering what is commonly talked or written about regarding a topic and making sense of those common elements.

TA is considered an umbrella term for numerous different ways of examining qualitative data [55], [59]. Both inductive and deductive approaches can be used with TA. An inductive approach to data coding and analysis is a bottoms-up approach, influenced directly by the data itself. This means that codes and themes are extracted from the data, ensuring that the analysis accurately matches the data's content. In contrast, deductive approach to data coding and analysis uses a top-down approach, using already pre-defined concepts, ideas and topics to guide coding and interpretation. This means that the codes and themes emerge from the concepts and ideas the researcher brings to the data, rather than from the content of data.

4.8.2 Affinity diagram

Thematic analysis is often done in group brainstorming using sticky notes [55]. Affinity diagrams is a commonly used technique for analyzing the data, uncovering themes and finding an overarching narrative. This technique aims to organize individual ideas and insights into a structured hierarchy that presents common themes and structures. The groups of which the notes end up in emerge from the data, rather than being predefined. The Affinity diagram is built progressively, starting with putting up one note on the board. Subsequently, the team searches for another note that in some way matches or is similar to the previous note, creating clusters of similar notes. Affinity diagrams can be conducted both physically with real sticky notes or using digital collaborative tools.

4.9 Tools

During the process several tools were used to facilitate the research and the communication. Google Drive platform served as a central hub to share articles and other research materials. Google Slides was used for updates to share with the supervisor at the university on a weekly basis. Google Docs was used to collect information and write notes and Overleaf was used for crafting the master thesis.

Microsoft Teams was used to conduct the expert interviews, which offered built-in recording and transcription features. For brainstorming and conducting the qualitative analysis of the results of the open-ended questions in the survey and the result of the expert interviews, tools such as Miro were used. Both Miro and Figma were used for creating figures and images.

Microsoft Forms was utilized to conduct the survey and to distribute them to the users. The data collected through the surveys was analyzed using Excel, and was used for the creation of various charts. Furthermore, IBM SPSS Statistics was used for applying more advanced statistical methods to the responses from the survey.

5
Planning

This chapter presents the initial timeplan of the study, shown in a Gantt chart. Further, it includes a table that lists the milestones that provides an overall description of the process and any changes that was made.

5.1 Initial timeplan

The execution of the project generally followed the initial timeplan, with only one week delay. However, the analysis phases of the qualitative and quantitative data took longer than expected, leading to overlapping activities than initially shown in Figure 5.1.

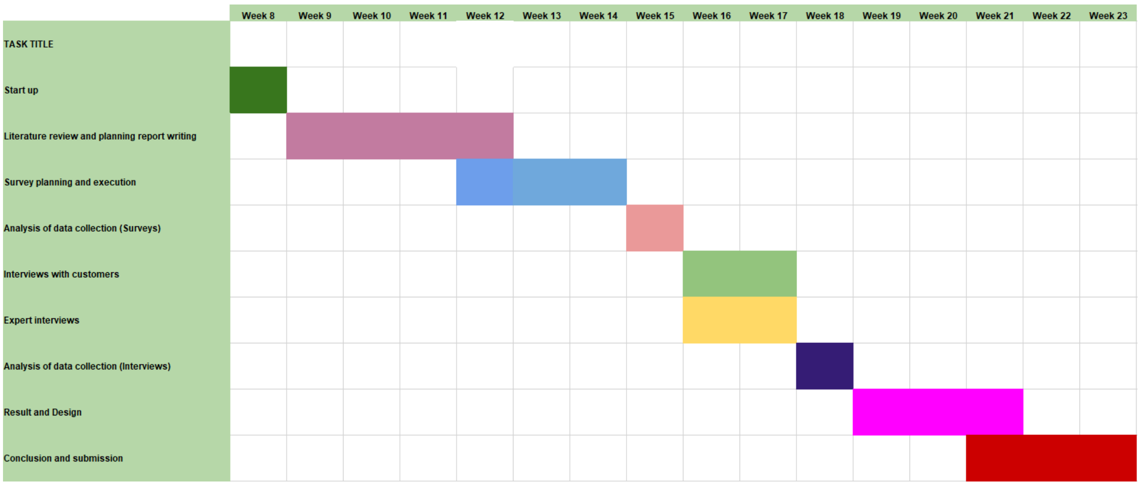


Figure 5.1: Gantt chart of initial timeplan

5.2 Milestones

Table 5.1 lists the initial project milestones. In the table, a 'X' corresponds to completed activity, while '-' marks an activity as incomplete. All milestones were completed except for the customer interviews, due to time limitations. The extensive data collected from the surveys could answer the first research question, and therefore expert interviews were prioritized over customer interviews. The expert interviews were crucial for exploring the second research question, which required specialized knowledge of the experts.

Activity	**Project Week**	**Completed**
Start up and preparing	8	X
Literature review and planning report writing	9-12	X
Survey planning and execution	12-14	X
Analysis of data collection (Surveys)	15	X
Interviews with customers	16-17	-
Expert interviews	16-17	X
Analysis of data collection (Interviews)	18	X
Results	19- 21	X
Conclusion and submission	21-23	X

Table 5.1: Status of milestones in the project

6

Execution

This chapter describes the execution of the methodological approaches utilized in this study. The execution consisted of several phases, beginning with defining the problem and reviewing the literature to understand the research. The next step was to formulate survey and expert interview questions, which consisted of several iterations for further improvement. After data collection, both quantitative and qualitative analysis were conducted. The quantitative was analyzed using statistical methods, while the qualitative data was analysed through a thematic analysis in six iterative phases.

6.1 Define problem

The study started with a few informal meetings with the supervisor at Volvo Cars, the aim was to gain a deeper understanding of the research area, current challenges and from what perspectives the problem could be investigated. Simultaneously as working closely together with our supervisor, exploratory research was done to investigate the latest trends, the state of other car manufacturers' voice assistants. At this stage we investigated many possible contexts to formulate our research within VAs in vehicles. Other topics explored were evaluating design principles for VAs in vehicles as well as investigating multimodal interaction in vehicles, combining a VA with an interactive display. However, given the advancements in generative AI may have a great impact on VAs, we decided to focus on this area. A Venn diagram was created to illustrate the intersection between the main areas: GenAI & LLMs, VUI and Automotive. The intersection between these areas showcases their common elements and highlights this studys focus area, see Figure 6.1.

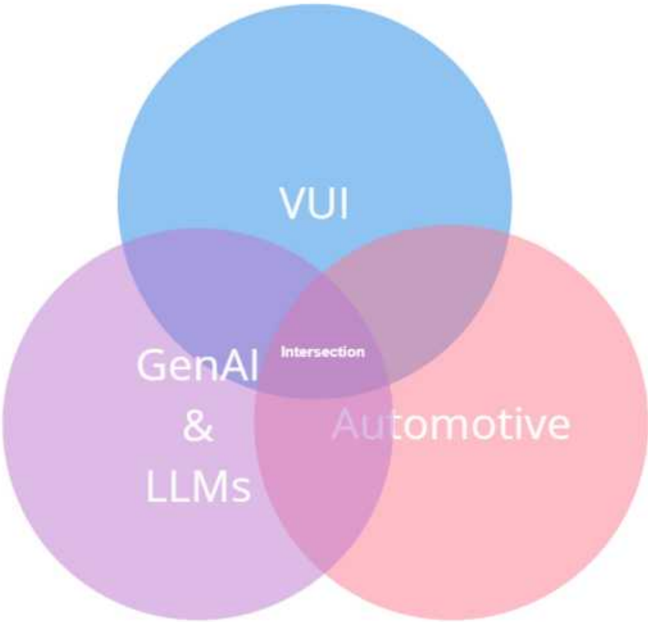


Figure 6.1: Venn diagram illustrating the research area of the study

6.2 Literature review

Once the scope of the study was defined a literature review was conducted, focusing on grasping the academic literature related to the topic of voice user interfaces and their technological aspects in the automotive domain. The literature search was primarily conducted through Google Scholar to discover relevant scientific articles on the topic, but also Institute of Electrical and Electronics Engineers (IEEE) and Human Computer Interaction (HCI) journals. In addition, to supplement these sources, relevant websites were explored. This strategy ensured covering the latest releases and developments in the field that might not have been captured in academic papers yet. Some of the keywords used are: *In-Car Voice Agents, Conversational Design, Generative AI, Conversational Agents, Voice Agents* and *LLMs*. These keywords could either be used separately or combined to get more precise results. A literature review was also conducted on the chosen framework, the Triptech method. Existing research and theories that were related to the Triptech method was examined to gather a comprehensive understanding of the methods relevance and utility.

6.3 Brainstorming survey questions

At this stage a brainstorming session was carried out where a few initial scenarios were constructed. The brainstorming session was carried out together with our supervisor at Volvo Cars. These scenarios were inspired from the Triptech method, where key aspects based on the Triptech method were followed to structure the scenarios. These are:

- **Non-technological user needs:** The questions should articulate clear, concise user needs without implying any specific technological solutions. This approach helps to measure the genuine interest and necessity of the features being considered.

- **First-person perspective:** Formulating questions from the user's viewpoint makes the survey more relatable and encourages respondents to consider their real-world experiences and needs.

- **Single need or problem per statement:** Ensures that the feedback you receive is specific and actionable, avoiding confusion about what aspect of the question is being addressed.

These three aspects were followed throughout the entire creation of the survey to maintain consistency and structure in the survey. The questions were posed as realistic situations of something happening while driving, such as the occurrence of an issue or problem, subsequently a suitable solution to address the issue at hand. The first of the three guidelines was only partially followed. The Triptech method has not been used in an automotive setting before and were therefore needed to be tailored to fit this study. As cars today have multimodal interfaces, such as combining a display with a VA, the scenarios and questions were adapted to better suit the automotive context and the focus of this study. The survey was framed in the context of a vehicle equipped with a VA, in this study it was called a "Car Co-Pilot". The purpose of naming it Co-Pilot was to build on the mental model that it is an assistant designed to support the driver in various situations. For each scenario two questions were asked 'How important is it to address this issue' and 'How likely is it that you would use the Car-Co pilot in this situation'.

To evaluate the user needs in the early stages of a design concept, the Triptech method utilizes a three distinct parameters to measure desirability. The parameters taken into account are scenario, frequency and importance, see Figure 4.1. These parameters have been adapted to better fit the purpose of this study. As Triptech originally measures the frequency of a problem or situation, this parameter has been substituted with the likeliness of using a VA for support in specific situations.

In the automotive context, the frequency of encountering issues or inconsistencies is relatively rare, so the frequency parameter would likely receive low scores. Instead, the likeliness of using a VA was used to address the user needs and desire for an in-car VA. The likeliness of using a VA to get assistance is measured against the importance of addressing the situation. This change is intended to measure the desirability of using a VA to get assistance or resolve a problem, see Figure 6.2

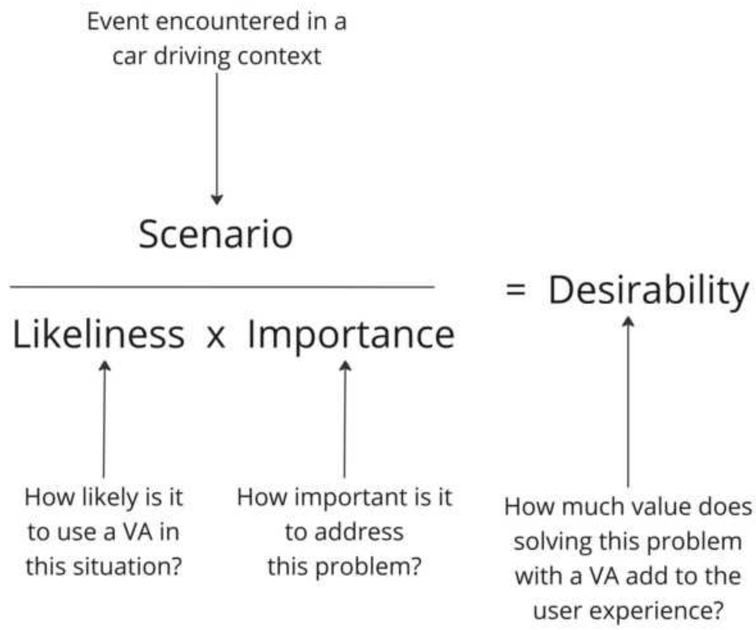


Figure 6.2: This study's approach of how to assess user needs in the early stages of product development

The brainstorming session was meant to explore different aspects of using a VA while driving and where it could be most useful. Figure 6.3 presents how the questions in the survey was tailored to suit the automotive context. This was improved throughout the entire process of crafting, iterating, and refining the survey questions. The figure is meant to give insight into how our method differs from the Triptech method.

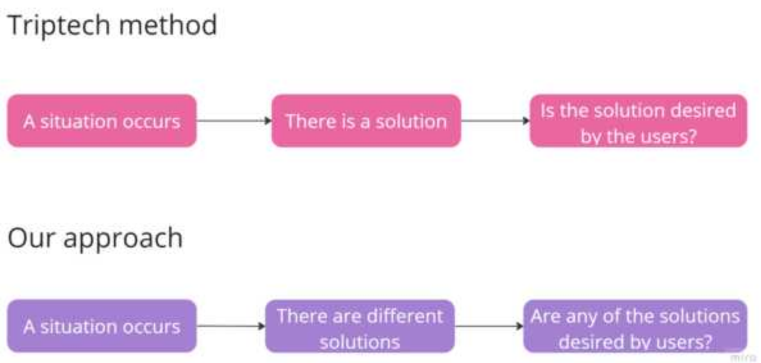


Figure 6.3: This study's approach of how the survey questions were tailored to suit the automotive context

The first aspect explored during the brainstorming focused on the five senses of the body and how a new issue or problem could be discovered. We crafted scenarios building upon the sense of the body that noticed the issue. The taste sense was not relevant in a car driving context and was therefore discarded, the focus was put on the senses of sight, sound, smell and touch. Scenarios were created based on problems that the user could encounter while in the car, starting with one or more of the senses. The survey included scenarios such as seeing a warning pop-up on the car's screen or hearing, feeling or smelling something unusual coming from the car.

The first draft consisted of approximately 15 questions. Figure 6.4 shows the iterative survey design process.

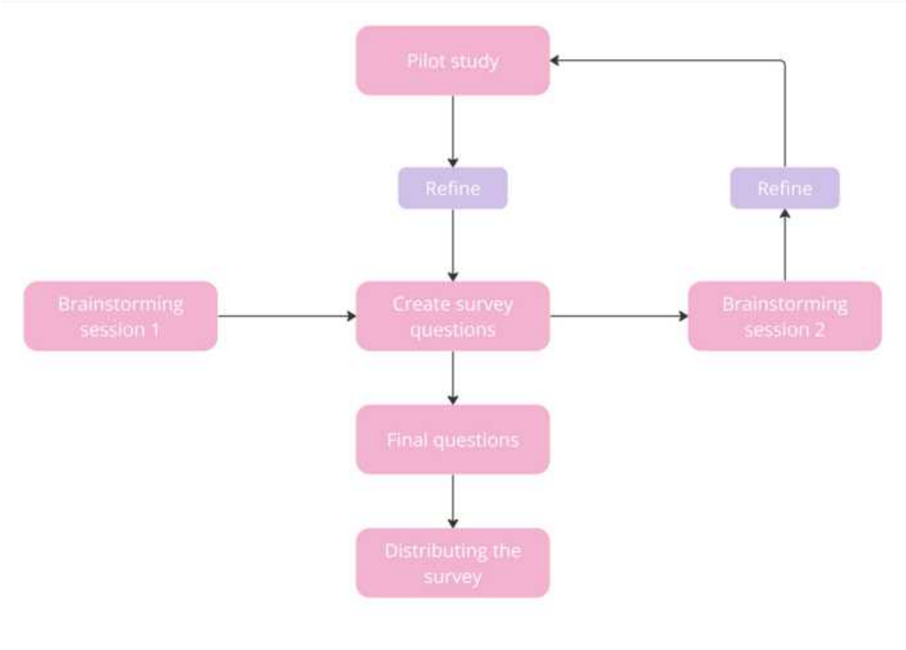


Figure 6.4: Diagram illustrating the iterative survey design process

6.3.1 Refining survey questions

After finishing the first draft of the survey questions a meeting was scheduled with the supervisor at Volvo Cars to review and refine the questions further. The brainstorming session aimed to finalize exactly what should be discovered from the responses from the survey. After a meticulous walk-through of all questions, it was concluded that the survey should provide realistic examples of situations in the car that the survey participants are already familiar with, such as an unusual sound coming from the car or having trouble finding how to perform a task. To further tailor the scenarios to the suit the automotive context, we propose three ways of resolving the issue or enhance the UX. As mentioned previously, each scenario starts with a situation or issue of some kind, such as: 'When I hear an unusual noise in my car', and then the following phrase proposes a solution to the issue. We formulated three different solutions as:

- I would like an **explanation** of what the problem is
- I would like to **report** the problem to Volvo
- I would like to know how to **solve** it

These are the three solutions that were were added to the scenarios of the survey. However, in a few scenarios only two of the solutions were suitable and therefore one of the solutions was removed. By adding this dimension of proposing different ways of resolving issues or handling situations, the study aimed to understand what assistance drivers need in different situations. This differs from the way Triptech formulated their scenarios as they only propose one tech agnostic and distinct solution to each scenario, such as When I hear music I like, I want to know the artist and title (e.g. a song playing in a coffee shop).

Subsequently, different perspectives, user needs and situations were explored to refine the formulated scenarios. The scenarios were organized into two main categories based on who initiated the interaction, the user or the VA. These two categories were subsequently broken down into subcategories.

6.3.1.1 User initiated interaction

This category is based on scenarios where the user is initiating an interaction with the voice assistant. The scenarios are divided based on the users initiation, see Figure 6.5. See Figure 6.6 for an example of a scenario question in the survey.

1. **Problem alert**

   These scenarios represent situations where the car notifies the user about something regarding the car. Notifications are usually shown due to sensors detecting some kind of issue or providing the user with information about an issue. The car gives the user visual cues, such as information on the display or dashboard of the car. All these scenarios had different levels of severity, low severity, medium severity and high severity. Severity levels were included to understand what kind of assistance the user wanted in relation to the severity of the scenario. See an example of a scenario with medium severity below. In situations when the car alerts or informs the user about something is relevant to this study as they could trigger the user to take action or get assistance utilizing a VA, the Car Co-Pilot.

   Example: When I see a warning pop-up on the screen (such as the parking sensors need to be cleaned or that the airbag on the passenger side has been deactivated)...

   ...I would like an **explanation** of what the warning pop-up indicates

2. **User perceived**

   This scenario represent a situation where the user notices something wrong with the car. The issue is noticed through the users senses, sound, smell or touch. In this scenario the car is not yet aware of the issue. This is usually known as rattles and squeaks and could come from various components of the car. For example, a squeaky noise may indicate that the break pads are wearing thin and need to be replaced. This question investigated if the driver would use VA to get assistance rather than visiting a repair service shop.

Example: When I notice something unusual with my car (such as odd noises, vibrations, or changes in my car's handling)...

...I would like to **report** this problem to Volvo

3. **User feedback**

   These scenarios represent situations where the driver may want to provide feedback about something regarding the car, share ideas and inform about an experience. Giving feedback includes situations where the user wants to report software issues or other aspects that are not working as expected. Sharing ideas investigates whether users want the ability to suggest new features or improvements. Informing about an experience covers situations where the user wants to praise a function they really like.

   These scenarios were relevant to the study as giving feedback, sharing ideas and informing about an experience could provide the manufacturer with valuable information about user needs and expectations. They also aimed to determine if users are interested in having a way to impact the development of the car they are driving. Allowing customers to influence car development could lead to a stronger community, which was also of interest to investigate.

   Unlike other scenarios where the user initiates contact with the VA, these scenarios focus only on feedback, sharing, and informing. It was not relevant to ask if the users would like an explanation, report, or solution. Although, reporting is similar to giving feedback, the question is asked differently for increased clarity to the survey participants.

   Example: When a new over the air (OTA) software update has been installed in my car...

   ...I would like to give **feedback** to Volvo about it

4. **Issue experienced by user**

   These scenarios represent situations where the driver encounters issues such as trouble in finding features, experiences unexpected responses, or has trouble understanding how a function works or how to use it, leading to an interaction with the VA to seek help. These scenarios were valuable in capturing the cognitive process of the users interacting with the car functionalities. The questions aimed to determine if there is a need for greeting support from a VA immediately and intuitively if the user encounters any potential cognitive load issues that may arise from complex systems. In two out of three scenarios, only two of the proposed solutions were relevant. The relevant solutions to each of these scenarios are shown in Figure 6.5

   Example: When I have trouble finding how to perform a task (such as setting up a new car profile or connecting my phone through Bluetooth)...

   ...I would like to know how to **solve** it (such as guidance)

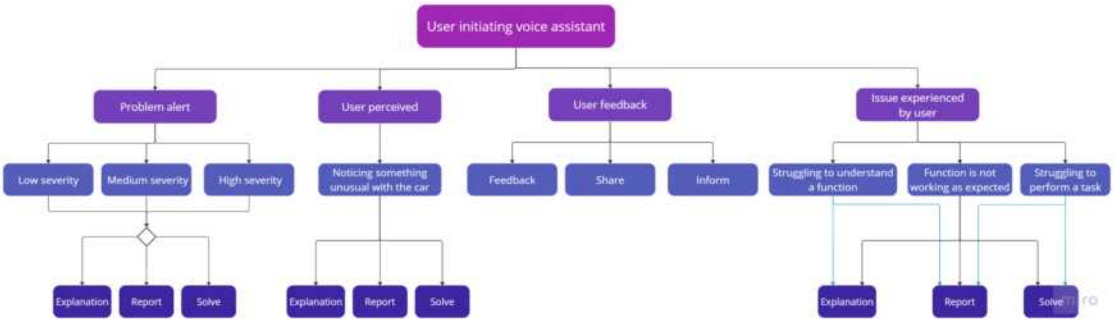


Figure 6.5: Structure of survey questions where user initiate an interaction with the VA

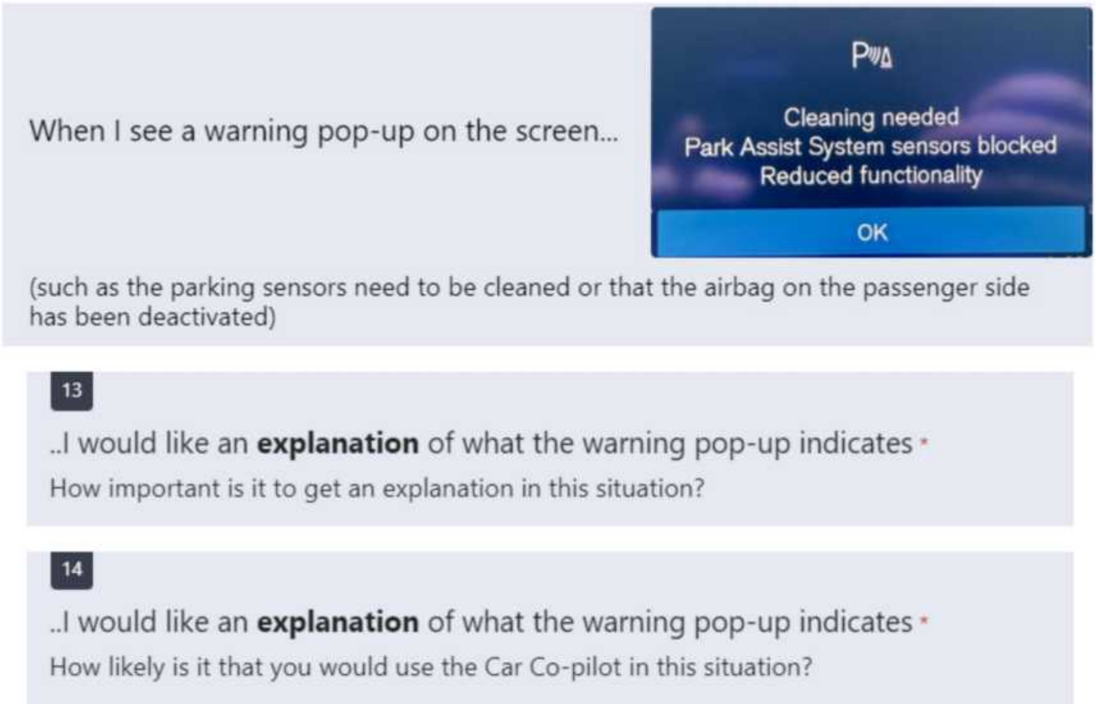


Figure 6.6: Example of a question in the survey

6.3.1.2 Voice Assistant initiated interaction

The scenarios presented in this category explore situations where the voice assistant proactivity initiates interactions with the user. These survey questions were included to determine if there was any interest in being contacted by a VA and the appropriateness. The goal was to see if the users wanted to interact with the system proactively, rather then only using it when they encounter problems or wants to provide feedback. Proactive interaction with users can be beneficial for the manufacturer, as it allows them to provide drivers with engaging information about driving, updates, and new features.

Example scenario: The Car Co-pilot asks me about my opinion on the new Over The Air (OTA) software update that has been installed in my car.

Each scenario aimed to evaluate the users view of how appropriate it is that the VA initiates the conversation in these situations.

6.3.2 Pilot study

When the first draft of the questionnaire was finished, a small pilot study was conducted with 7 people. The pilot study was important to get a grasp of the duration of the survey and ensure that participants understood the questions posed. After reviewing the feedback from participants in the pilot study, the questionnaire was refined and then sent to the person responsible for the distribution at the CoDev department. The feedback from the pilot study indicated that the survey was too vague. As the survey was highly hypothetical it needed specific examples to clarify the referenced scenarios. Accordingly, for each scenario where 'get help solving the issue was the solution, phrases like 'such as troubleshooting or guidance' were added for clarity. Relatable examples were included in each scenario to make it easier for the participant to evaluate the scenario by referring to something familiar. In the first scenario, the question 'When a warning pop-up shows up on the screen of my car', was followed by an example in small text: 'such as the parking sensors need to be cleaned or that the airbag on the passenger side has been deactivated' to make the question more relatable.

Also, as mentioned previously, the statements should remain tech-agnostic to be able to measure a genuine interest and needs for the features. However, despite this suggestion, this survey did not strictly follow this guideline due to the survey becoming too hypothetical. Therefore, an informational section was added before asking the situational questions, to ensure users understood the questions correctly. The information below is what was presented to the users.

'Imagine having an in-car voice assistant 'Car Co-pilot' that knows about your car, can give support and help you report problems. In the following section, think of the questions as if you had this 'Car Co-pilot' voice assistant in your vehicle.

You will be presented with a variety of scenarios related to your experiences while driving. Each scenario is designed to explore different aspects of using a Car Co-pilot system.'

Definitions were also added to the questions about trustworthiness and reliability, as it can be difficult to differentiate between the two. Additionally, the text in the solution part of the question, 'explain, report, solve' was refined, aiming to make the survey more consistent and use similar wording across different solutions. Furthermore, the rating scale for importance and likeliness were different. Importance was measured using the scale not at all important' to 'extremely important', while for measuring likeliness the scale 'very unlikely' to 'very likely' was used. To maintain consistency in meaning, the scale was changed to 'not at all likely' to 'extremely likely', in all scenarios. This rating scale was chosen as it is the scale used in the Triptech method. Lastly, the introductory text in all sections was shortened and clarified, and the overall language throughout the survey was refined to enhance clarity and consistency.

6.3.3 Distributing the survey

The questionnaire was created online using Microsoft Forms which was a convenient method to efficiently gather large samples of data and simplify the analysis process later on [52]. The CoDev department uses Microsoft Forms for distributing forms and questionnaires to their CoDev-drivers. Given that this survey is conducted in collaboration with the CoDev department, it was a requirement of using Microsoft Forms for the questionnaire. The survey was sent out to a total of 400 people, 300 Swedish drivers and 100 American drivers. The difference in distribution is due to the significantly lower number of American Co-Dev drivers. The survey was crafted in English and all participants were given the same survey. Appendix A for the final questions in the survey.

6.4 Expert interviews

Expert interviews were conducted to determine whether experts in the field possess any information on the potential improvements of in-car VAs through GenAI. The interview also aimed to gain additional insights into user needs and the specific functionality drivers are seeking, to complement the findings from the survey. To ensure consistency between the survey responses and the expert interview responses, all expert interviews were conducted in English.

6.4.1 Preparing the questions

When constructing the questions for the interviews, focus topics and areas were clarified to ensure responses aligned with the research questions and goals of the interview. The interview questions were open-ended questions and gave flexibility for follow-up questions and further discussion. Moreover, some of the questions were similar but rephrased in another way to capture more nuanced answers.

Through Volvo Cars, three experts were made available for interviews, all in the field of voice assistants but possessing individual expertise. The first interviewee is specialized in computational linguistics and speech technology. Working with digital assistants such as telephone dialogues, conversational systems, practically and test systems. The second interviewee is specialized with voice, digital assistants and customer needs and is well-acquainted with ecosystems for digital assistants and competitors. The third interviewee is specialized in the development of dialogues for vehicles and works with integrating Google Assistant and collaborates with Google on voice and dialogue systems in cars. With their specific expertise area in mind, some of the questions were tailored to their respective fields.

6.4.2 Executing the interviews

The interviews were conducted online via Microsoft Teams to enable the softwares built-in features, the recording and transcription. These features simplified the process of compiling the data afterwards. In the beginning of each interview, the participants gave their consent for recording and transcription. There were in total three interviews, each on separate occasions. Every interview was designed with semi-structured questions and lasted for approximately 30 minutes. Each interview session consisted of one interviewee, one interviewer and one note-taker, which provided a balanced and effective setup for collecting comprehensive data. The note-takers task was to complement the transcription and the recording by documenting the key observations. The document was shared in Google Drive after the interviews.

6.5 Analysis of quantitative data

The analysis of the quantitative data in the survey was divided into separate parts. Initially, the responses from the survey were imported into a table in Excel from Microsoft Forms. To facilitate the analysis process, the cells with textual rankings were converted into numerical values as follows:

- Not at all important/likely = 1
- Somewhat important/likely = 2
- Moderately important/likely = 3
- Very important/likely = 4
- Extremely important/likely = 5

This part of the analysis aimed to compare the importance of handling a problem and likelihood of using a Car Co-pilot system in various scenarios. Inspired by the Triptech method the purpose was to determine the percentage of how many rated the scenario problems as Very important/likely (4) or Extremely important/likely (5). This was calculated by using Excels COUNTIF function to count respondents rating 4 or 5. The percentage of these responses were then calculated using the sum of 4s and 5s divided on the total number of responses for each question.

The data were then illustrated using Scatter Plots, inspired by the Triptech method to identify which problems to address first. A Scatter Plot graph, presented in section 7.1.1, was created for each dimension of the problems: explanation, report and solve (and the same structure for user feedback). Each graph plotted the likeliness of using the Car Co-pilot (y-axis) against the importance of the problem or situation (x-axis) in these dimensions. The graphs were highlighted in blue to mark the area where there was an intersection between the importance of the problem and the likeliness of using a VA of the highest rated survey questions. The analysis of this data was conducted with the combined responses from Sweden and the United States, as well as separately for each nation to allow for comparative evaluation.

Moreover, a chart, presented in Figure 7.14, representing the mode and median was created to analyze the responses regarding the Car Co-pilot initiating conversations. The mode and the median were calculated using Excel's built-in MODE and MEDIAN function of the desired data columns. Incorporating both mode and median into the analysis helped to capture the most typical and the most frequent responses, and thereby provided a more comprehensive understanding of the data.

Finally, in Excel the percentage of the screening questions as well as the user demand of using a Car Co-pilot were calculated. The analysis was conducted using the COUNTIF function for each possible response option. The percentage of these responses were then calculated using the sum of each responding for each question divided on the total number of responses.

The screening questions cover aspects such as the gender, age, familiarity of voice assistants and conversational chatbots, what they are used for, and the perception of their reliability, trustworthiness and usefulness. The result for the user demand of using a Car Co-pilot was presented in a pie chart, presented in Figure 7.15.

6.5.1 Spearmans rank correlation

To determine any potential correlations among the responses to the survey, a correlation matrix with all questions included (except gender, age, nationality and the open-ended questions) was created in IBM SPSS Statistics. IBM SPSS is a software tool mainly used for statistical analysis. It has a built-in function to compute various types of correlations. In this study, the Spearmans rank correlations were employed. IBM SPSS automatically adjusts for tied ranks which ensures accurate calculations [60].

After the matrix (Figure 7.20) was created in IBM SPSS it was imported into Excel for refinement. The correlation matrix is symmetric around its main diagonal, where values above the diagonal mirror those below. To enhance the readability of the matrix, the upper half of the matrix was removed in Excel. Furthermore, color coding was used to differentiate the strengths of correlations, which is a technique to improve the visual clarity and accessibility, which aids to make it easier to quicker spot significant relationships within complex data.

6.5.2 Mann-Whitney U test

To determine any statistical significance between the two chosen groups, i.e. Sweden and the United States, a Mann-Whitney U test was performed in IBM SPSS. This is a built-in function within the software to enable these types of analyses.

The test analyzed the responses to all the questions in the survey (except gender, age, nationality and the open-ended questions). The result was compiled into a table, which was later imported into Excel for refinement. Furthermore, color coding was used to highlight the values indicating significant differences between the two groups, particularly those equal to or below the predefined significant level.

6.6 Analysis of qualitative data

The qualitative data analysis was conducted using a Thematic Analysis in six steps. At first, the analysis was conducted separately for the survey and interview questions as these were seized at different times. Once the data from both the survey and the interview were gathered these were analyzed together as an entire dataset. The survey data was captured in an excel sheet while the interview data was captured through notes and transcriptions from the interviews. All data pieces were later transferred to Miro as sticky notes.

6.6.1 Phase 1: Familiarizing with the data

An initial pass though the data was conducted by both researchers to become familiar with the content. This examination included reading and re-reading all textual data such as notes, transcriptions, and textual responses from the survey. Additionally, each interview was listened to an extra time to ensure no details were missed. All survey responses were pasted into the collaborative tool Miro and displayed as sticky notes to enhance readability and facilitate analysis.

During the review process, notes were taken simultaneously to capture insights as they emerged. The immersive and iterative approach to reading, listening and taking notes was essential in gaining a deeper understanding of the collected data. This process went beyond just reading the words, it involved trying to understand the underlying tone and meanings conveyed by the respondents. This method not only helped in comprehending the nuances of the data but also in identifying and highlighting early points of interest that could be significant for further analysis. The notes at this stage were of an observational and casual kind, serving as personal reminders. Their informal character made them helpful for remembering impressions during the deeper analytical phases that followed.

6.6.2 Phase 2: Generating initial codes

In the second phase, a systematic analysis of the data was conducted. The objective of this phase was to create the first building blocks in out thematic analysis, by generating initial codes. The codes were developed by looking at the characteristics of the data and assigning labels to features that may be relevant to the research question. The codes derived from the survey data were generated on a semantic level of meaning, indicating that they are highly descriptive and closely aligned with the actual content of the data. Concurrently, the codes from the interview data were primarily generated on a semantic level of meaning as well, however, given the complex nature of the interview responses, some codes were on the latent level of meaning, capturing meaning that lie beneath the surface of the data. Some of the codes for data collected in both the survey and interviews had a mix of descriptive and interpretive level of meaning.

During this phase, the entire data content was encoded to identify all potentially relevant aspects related to the research questions. The coding process included not just a literal transcription of the data but also a nuanced interpretation. For example, certain codes mirrored what the participants' said directly, while other parts of the data were interpreted beyond the explicit language of participants. This dual approach facilitated a more thorough understanding of the data, blending descriptive coding with interpretative analysis.

The coding was done using the collaborative digital tool Miro with both researchers actively analyzing the data. Each piece of data was analyzed fully before moving on to the next one, with each potentially relevant aspect being coded. Sometimes a single piece of data was coded in various ways to ensure all relevant aspects were encoded. The coding was done iteratively in an iterative manner where initial codes were frequently updated and expanded as new data was analyzed, creating clusters of data with similar characteristics.

When the data were completely coded and the data relevant to each code had been collated the second phase was ended. Resulting in a diverse set of codes, ranging from straightforward descriptions to more nuanced interpretations, acted as a foundation for the subsequent phases of thematic analysis.

6.6.3 Phase 3: searching for themes

In the third phase, the focus shifted from initial codes to constructing themes, crucial for capturing important aspects of the data in relation to the research question. As themes are not simply discovered this transition required another review of the data. By revising the coded data, identifying similarities, and clustering data that had common features, meaningful themes were formed.

During this phase, a thorough analysis of the previously formed codes related to broader topics such as Ease of use and Safe interface were clustered. These clusters helped in forming themes that reflected the participants' experiences and opinions regarding these topics.

The process also included creating thematic maps to represent the relationships and overlaps between themes, making sure that each theme contributed to a nuanced understanding of the data. This way of representing the data helped in aligning the themes into an overarching narrative of the research and deciding how they are connected to form a cohesive story. To conclude the phase, all data pieces relevant to the identified themes were clustered to be prepared for a subsequent revision phase. The transition from codes to initial themes is illustrated in Figures 6.7, 6.8, 6.9, 6.10 in phase 5.

6.6.4 Phase 4: Reviewing potential themes

In the fourth phase a recursive analysis of the data was conducted. The identified themes were reviewed in relation to the coded data and the dataset to make sure that the themes were aligned with them. To aid this process a few questions inspired from Braun and Clarke [59] were asked for each of the themes. The questions were formulated as following.

Could this be considered a theme or is it merely a code? If so:

- How effectively does this theme contribute to understanding the dataset and addressing the research question?
- What are the limits of this theme, what is included and what is not?
- Is there enough meaningful data to support this theme? Is it robust or superficial?
- Is the data within this theme too varied? Does it lack coherence?

Once a distinctive and consistent set of themes were developed, the process of reviewing the themes in relation to the entire dataset took place. This included a final reread of each data piece to make sure the themes capture the entire dataset or the aspects relevant to the research question. At this stage, only small refinements were done to the themes.

6.6.5 Phase 5: Defining and naming themes

The fifth phase of the thematic analysis focused on defining and refining the identified themes so that they are clear, focused and aligned with the research question. Each theme is clearly defined with a precise scope and purpose, to make sure they are unique and collectively contribute to a coherent story. This phase included selecting parts of the data that exemplified each theme. The selected data pieces were meant to demonstrate the breadth of the theme, giving a sense of what is included. The quoted and analyzed data pieces provided the structure for the analysis, giving the reader an understanding of how the data was interpreted. The themes were also categorized into three distinct groups, Challenges, Opportunities, and Functionality. These categories emerged from the initial themes to clarify and provide relevance to our research question.

Subsequently, the naming of themes were reiterated to make sure that they conveyed the right message, aiming for names that were informative, concise and catchy. By ensuring that themes and their names were clearly defined and supported by selected data, this phase laid the foundation for the final analysis and presentation of results. The final themes with their refined names are shown in Figures 6.7, 6.8, 6.9 and 6.10 below. The themes are also divided into several categories which are Capabilities, Limitations, Opportunities, User Needs, Challenges. These will be further presented in Chapter 7.

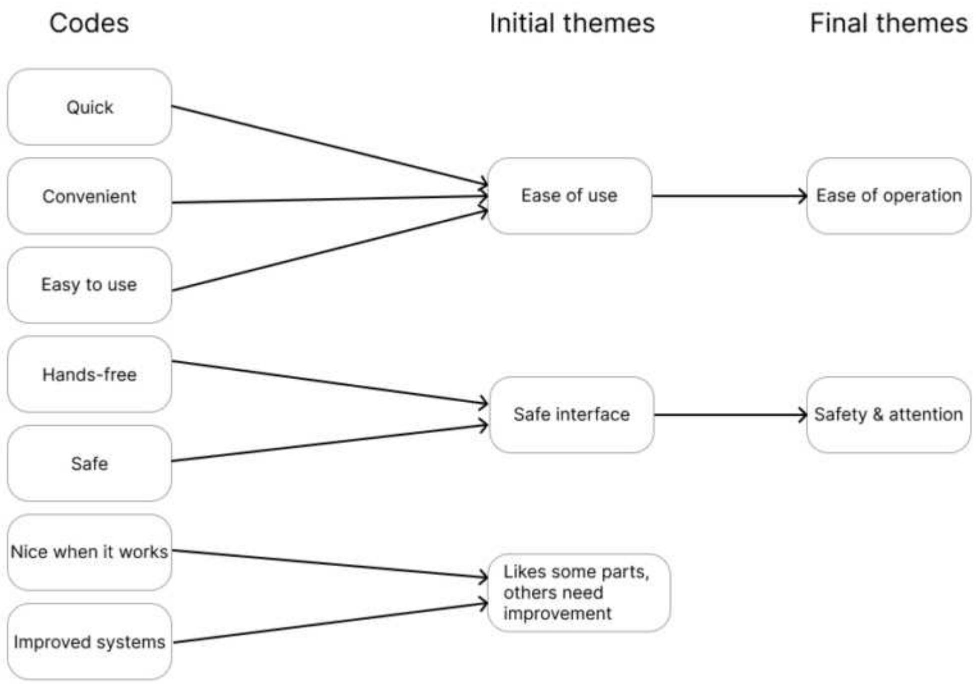


Figure 6.7: Thematic Analysis of VAs: Identifying themes and codes related to capabilities

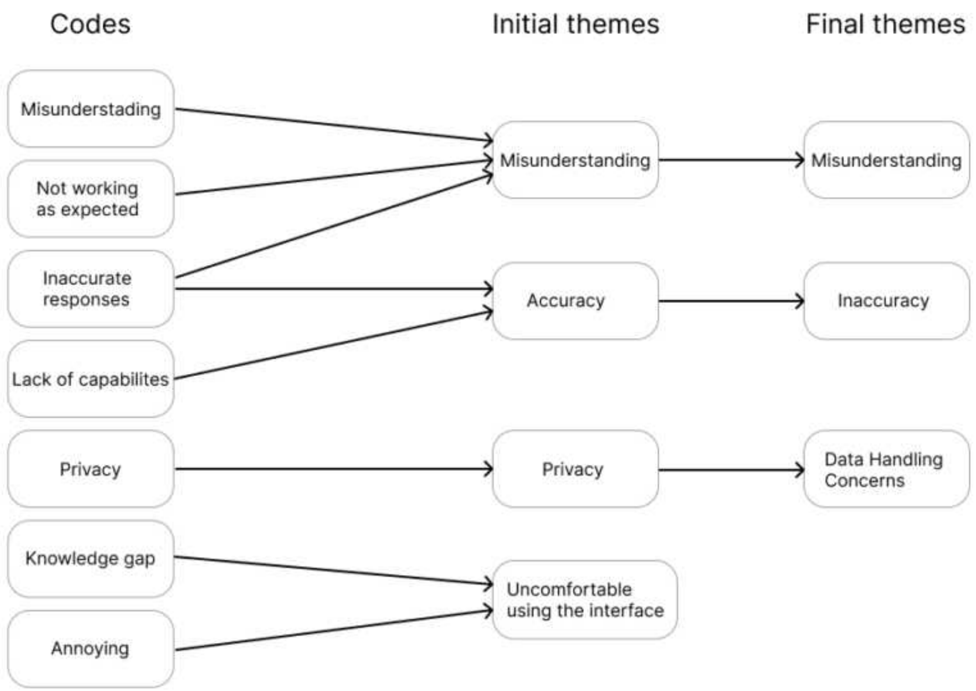


Figure 6.8: Thematic analysis of VAs: Identifying themes and codes related to limitations and challenges

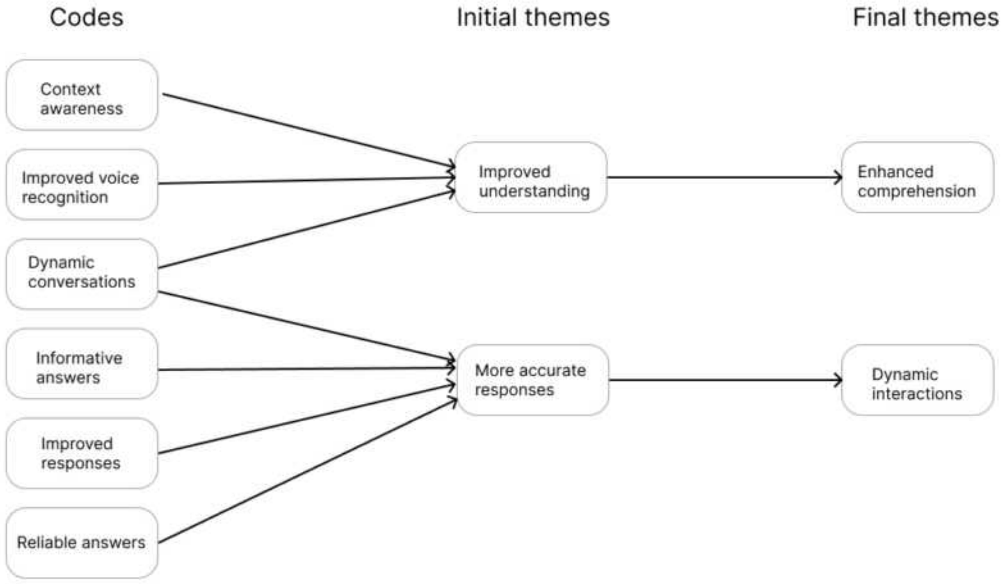


Figure 6.9: Thematic analysis of VAs: Identifying themes and codes related to the potential of generative AI

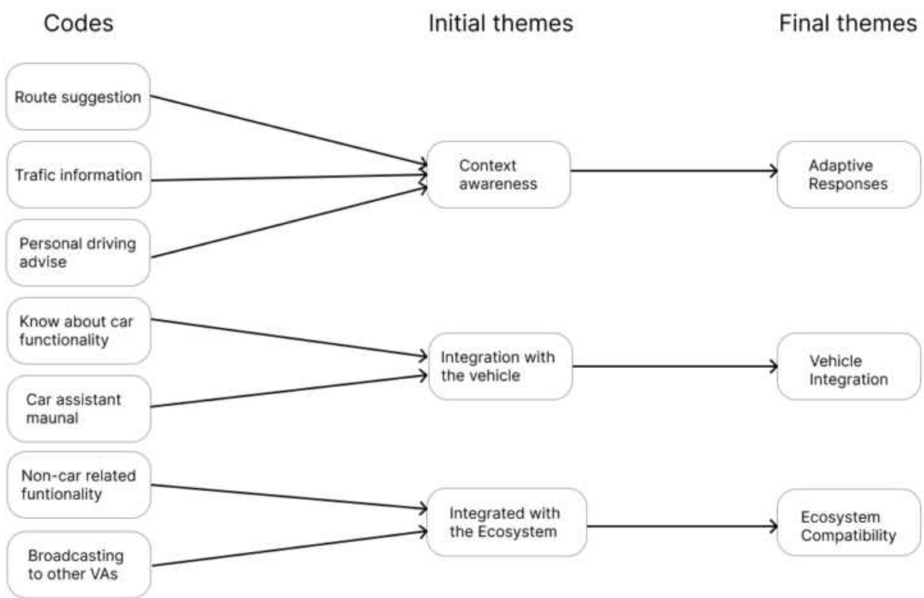


Figure 6.10: Thematic analysis of VAs: Identifying themes and codes related to user needs and functionalities

6.6.6 Phase 6: Producing the thesis

In the final phase of the thematic analysis, the focus was on producing the master thesis. This phase did not begin after all other phases were completed, rather it was interwoven throughout the entire research process. This includes everything from the informal notes and memos to the formal writing of the analysis and thesis. The objective was to present the findings from our data clearly, distinctly and authentically. Writing and analysis were two interconnected activities, each influencing and enhancing the other, as the thesis aimed to go beyond describing the data. Instead, interpret it in a way that addresses the research question. This iterative integration helped refine the narrative, ensuring the thesis avoided repetition and unnecessary complexity.

7

Results

This chapter present the result of the quantitative data of the survey responses followed by the qualitative findings from the survey's open-ended questions combined with the responses from the expert interviews.

7.1 Quantitative data findings

The survey was sent to 400 Volvo Cars' CoDev drivers, with 300 to Sweden and 100 to the United States. In total, 171 responses were received, including 139 from Sweden and 32 from the United States. Figure 7.1, 7.2 and 7.3 presents the background information of the responses. In Appendix A, the questions from the survey is presented.

Figure 7.1: The gender of the participants

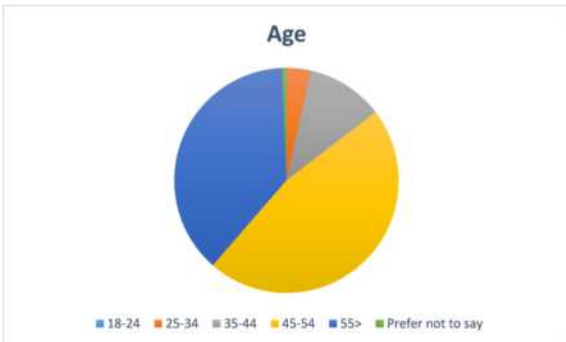


Figure 7.2: The ages of the participants

Figure 7.3: The nationalities of the participants

Figure 7.4 displays the percentage distribution of how familiar the respondents are with voice assistants, such as Google Assistant, Amazon, Alexa or Siri. Figure 7.5 illustrates the respondent's familiarity with the latest conversational chatbots, such as ChatGPT, Bing, Microsoft Copilot or Gemini. Participants who indicated any usage with these technologies were asked to specify their mainly uses. The results in the following section will only describe the percentages of respondents who were asked to answer these questions.

Results showed that 79% of the respondents commonly use voice assistants for navigation and traffic, such as setting destinations and getting traffic updates. Approximately 70% reported that they use it for hands-free communication such as making calls or sending messages. Around 44% reported using it to voice-control media and 36% use it for voice controlling the vehicle. All other reported usage areas fell below 22%

Furthermore, the most common use of conversational chatbots, reported by 75%, was asking a variety of questions to search for information, followed by getting assistance in writing and editing which was reported by 50% respondents. Approximately 35% claimed using them for educational and learning purposes. All other reported usage areas fell below 15%. Moreover, respondents who were familiar with conversational chatbots, including those who never used it but heard about it, see Figure 7.5, were asked to rate the reliability, trustworthiness and usability of this technology. The result showed that the most common rating (mode) was equal to 3 on a Likert scale ranging from 1 to 5 for each individual question.

Where:

- Unreliable/Untrustworthy/Useless = 1
- Reliable/Trustworthy/Useful = 5

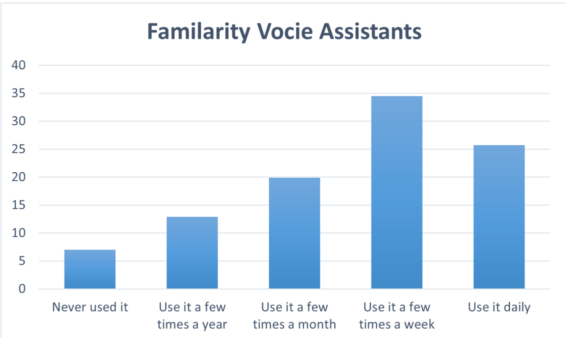


Figure 7.4: Participant's familiarity with voice assistants

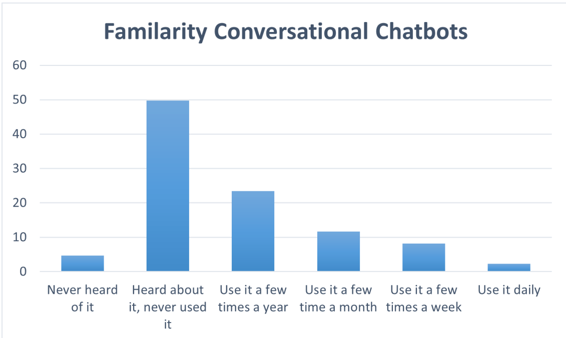


Figure 7.5: Participant's familiarity with conversational chatbots

7.1.1 User initiating interaction

The second part of the survey presented various scenarios related to the user's experience while driving. The scenario questions in this section are identified in the Scatter plots by a label that includes 'Q' followed by a number. For additional information on which plot corresponds to which question in the survey, see Appendix

B.

Figures 7.6, 7.7, 7.8 and 7.9 illustrate Scatter plots presenting the responses of those who rated a problem very- or extremely important (x-axis) against very- or extremely likely to use a VA (y-axis) in a certain situation. These Scatter plots present the percentages of the responses, including responses from both Sweden and the United States. The plots in the highlighted areas identifies situations where the user's needs are greater.

Figures 7.6, 7.7, and 7.8 look at specific key solutions for the majority of the scenarios in the survey. The key solutions refers in what way the voice assistant in the vehicle could offering support to the driver, e.g., explanation, report, or solve.

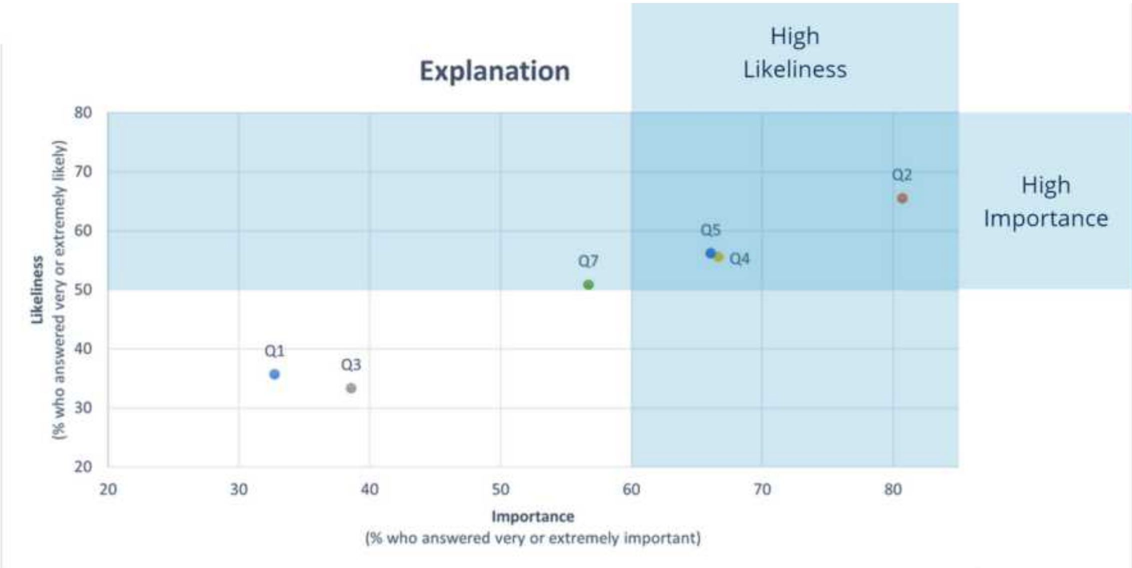


Figure 7.6: Participant's need of receiving an explanation in different scenario questions

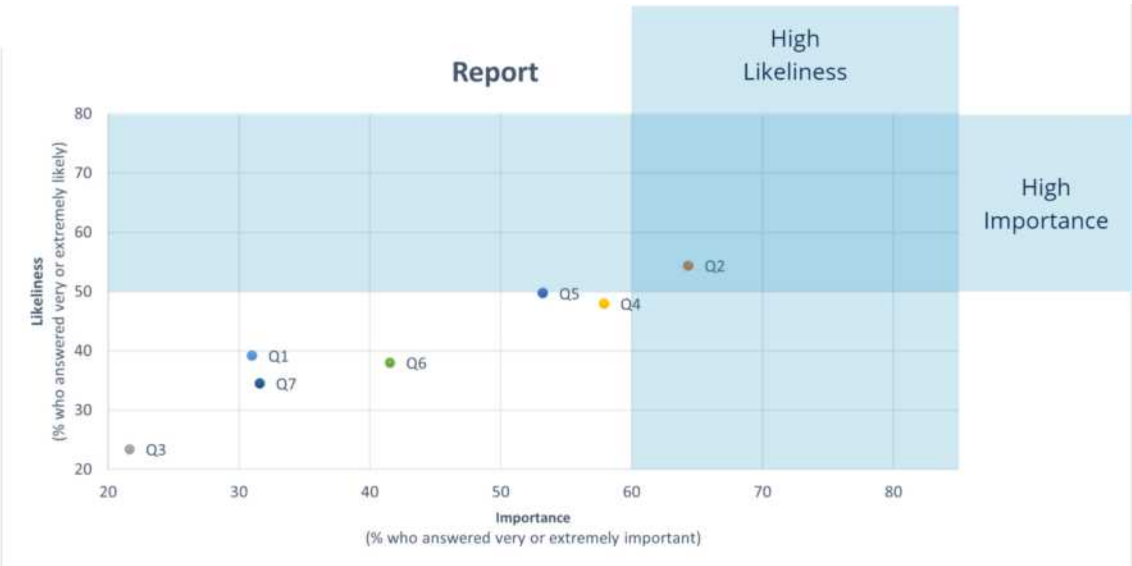


Figure 7.7: Participant's need of reporting in different scenario questions

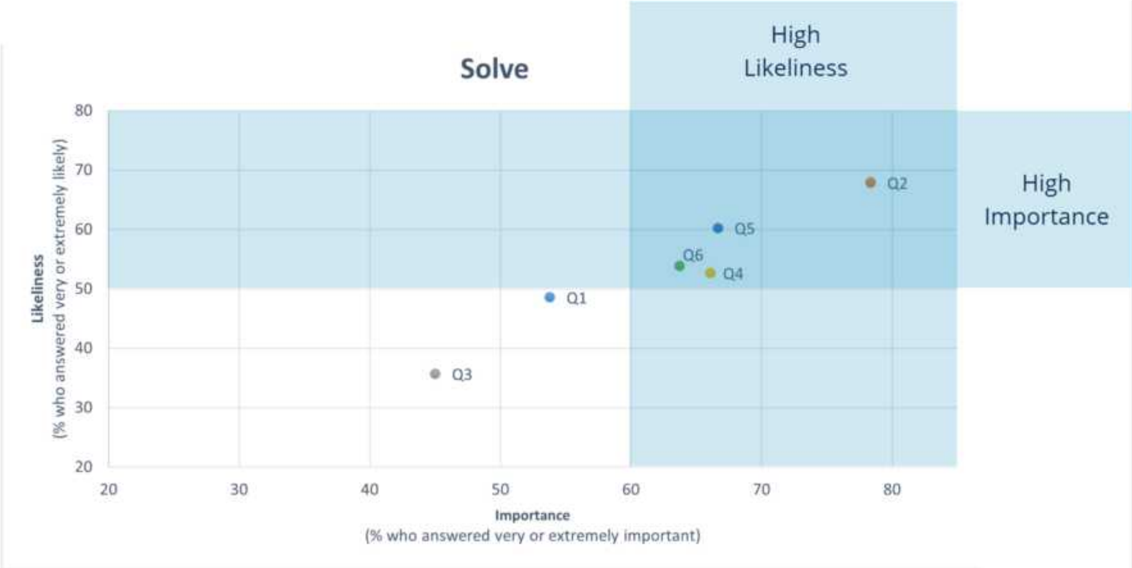


Figure 7.8: Participant's need of receiving support to solve a problem in different scenario questions

Figure 7.9 presents data from three specific scenarios where the user initiates conversation by providing feedback, informing, or sharing anything with the VA.

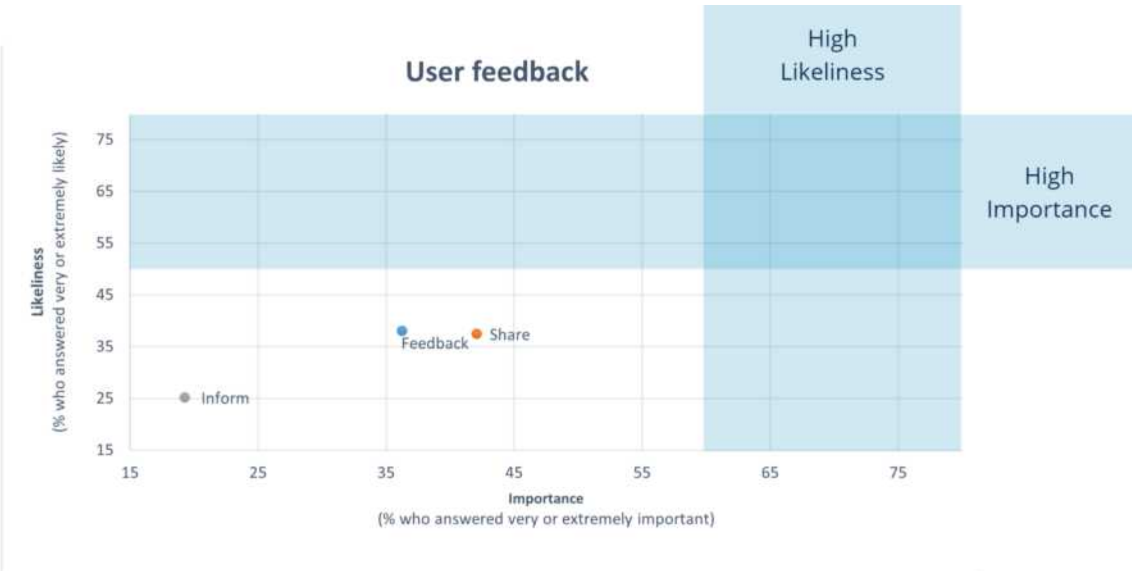


Figure 7.9: Participant's need of giving feedback in different scenario questions

Figures 7.10, 7.11, 7.12, and 7.13 follow the same Scatter plot format as 7.6, 7.7, 7.8, and 7.9, but they distinguish and compare the responses between the two nationalities, the United States and Sweden. In these figures, the from the United States are marked with a red star, while responses from Sweden are marked with a blue dot.

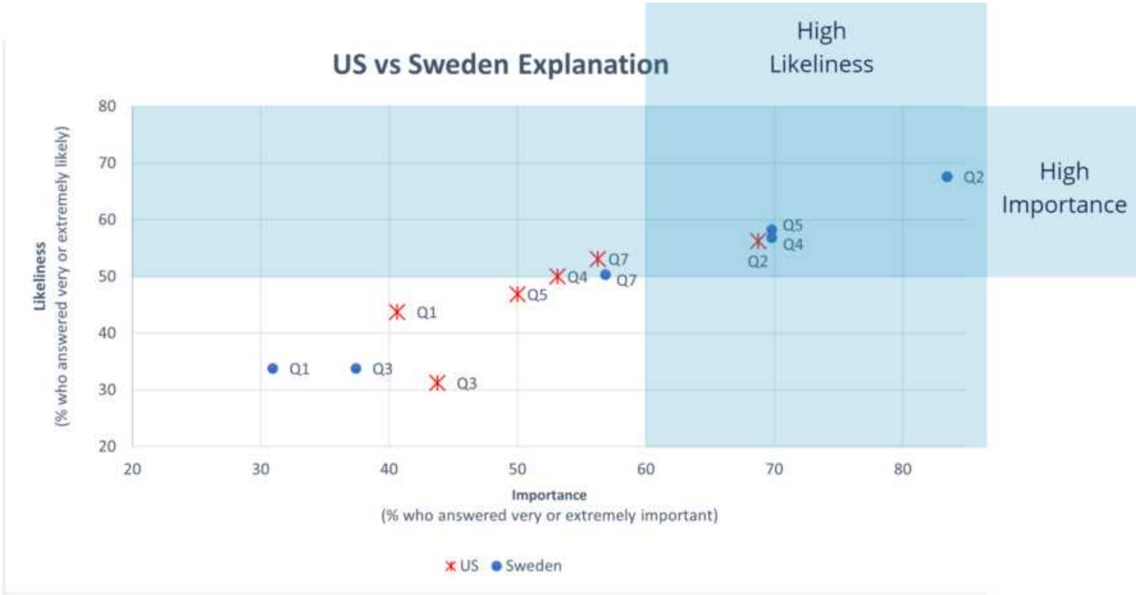


Figure 7.10: Participant's need of receiving an explanation in different scenario questions, comparing US & Sweden

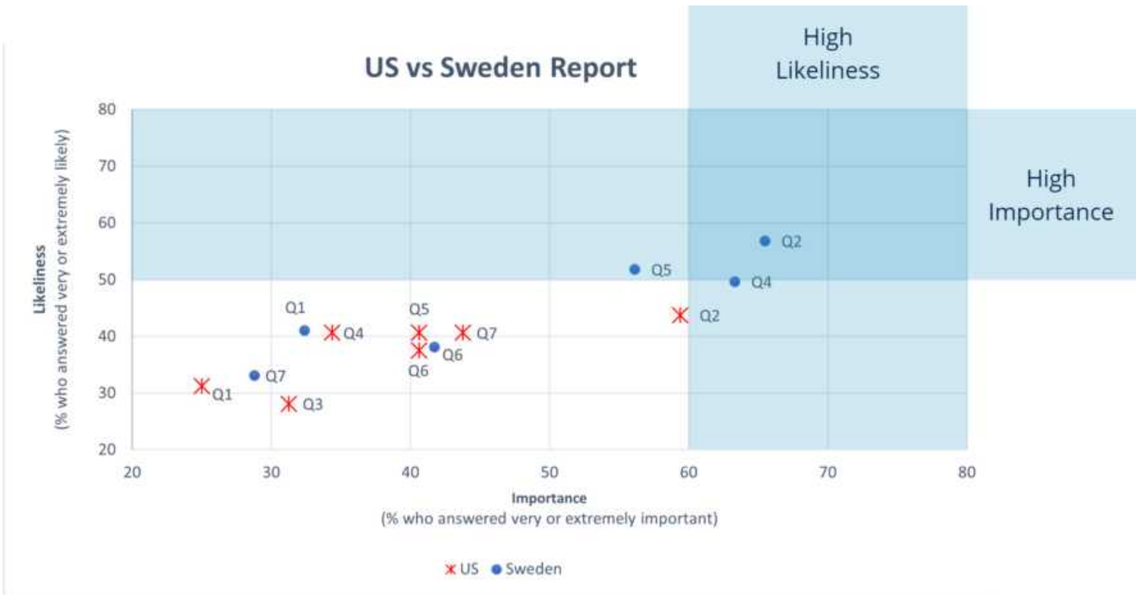


Figure 7.11: Participant's need of reporting in different scenario questions, comparing US & Sweden

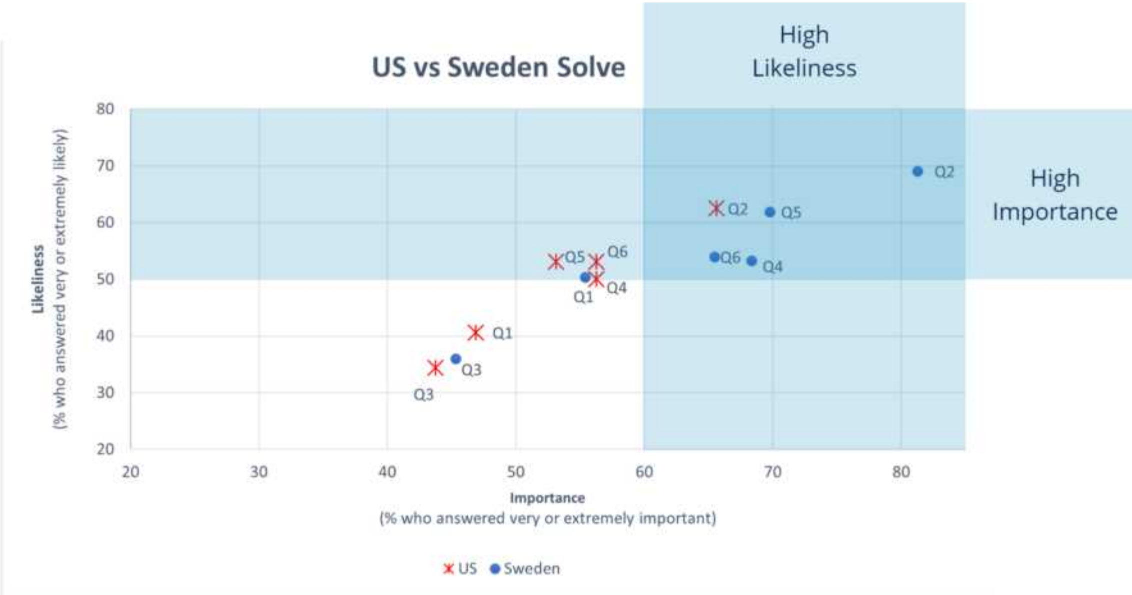


Figure 7.12: Participant's need of receiving support to solve a problem in different scenario questions, comparing US & Sweden

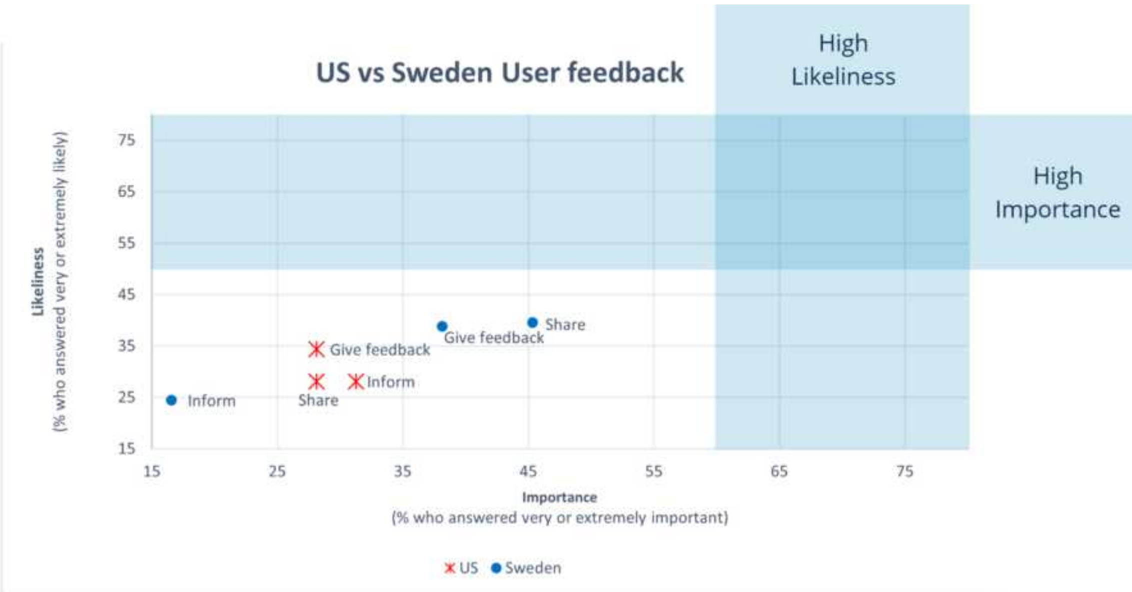


Figure 7.13: Participant's need of giving feedback in different scenario questions, comparing US & Sweden

7.1.2 Voice Assistant initiated interaction

The third part of the survey presented three scenarios of where the VA is initiating the conversation. The respondents were asked to rate the appositeness of using the VA (Car Co-pilot) for each statement. The scale was between 1 to 5 where 1 equals 'Not Appropriate' and 5 equals 'Very Appropriate' Figure 7.14 illustrates the median and mode of the respondent's ratings for these scenario questions. The orange bar presents the median values and the blue bar presents the mode. Results shows that median and mode are 4 for the first two scenario questions and 3 for the last one.

These ratings presents the numbers of the responses, including responses from both Sweden and the United States.

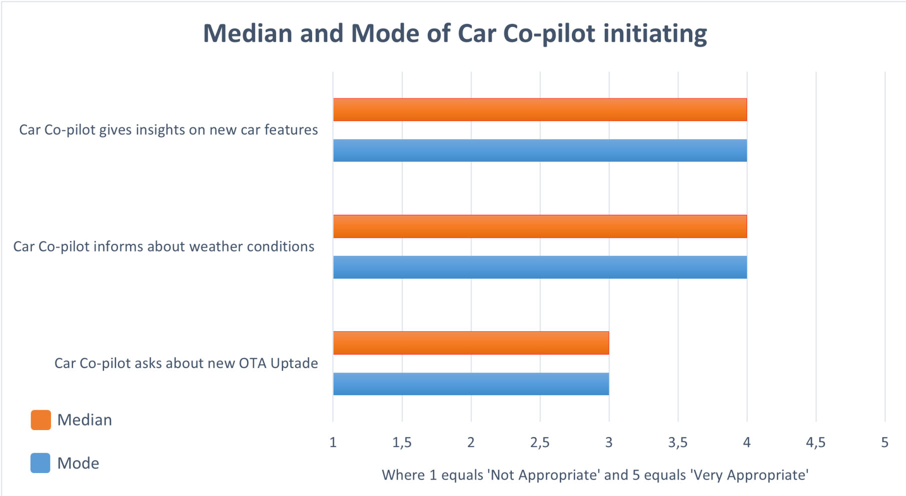


Figure 7.14: User need of VA initiated interaction in different scenarios

7.1.3 User Perceptions of In-Car Voice Assistants

In the last part of the survey, respondents were asked to provide additional thoughts about using an in-car voice assistant. The purpose of these questions was to gain a deeper and a more nuanced understanding of their perception of such technology. Figure 7.15 shows a pie chart of how likely the respondents would use a VA, like a Car Co-pilot to give more feedback, ask questions or report problems. It displays the combined responses from Sweden and the United States, showing that 56% of the survey participants are very likely or extremely likely to use VAs.

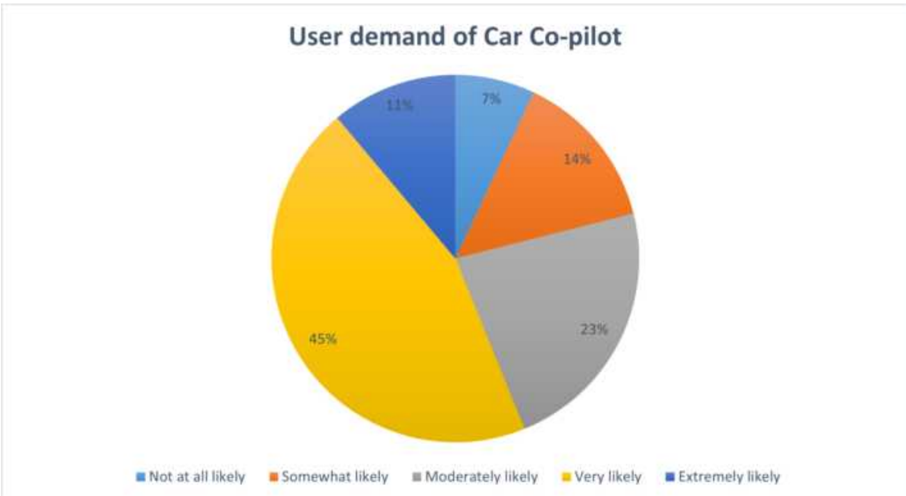


Figure 7.15: User likeliness of utilizing In-Car VAs to give more feedback, ask questions or report problems

7.1.4 Inferential statistics

This section presents the statistical findings from the Mann-Whitney U test and Spearman's rank correlation.

7.1.4.1 Regional differences

The Mann-Whitney U test showed a significant difference between the United States and Sweden in four questions asked in the survey. A p-value at ≤ 0.05 indicates that the observed differences are statically significant, meaning that they are unlikely ($<$ 5%) to have occurred by chance.

The familiarity of using voice assistants was compared between Sweden and the United States. On average, Sweden (Mdn = 4) was equal to the United States (Mdn = 4). However, a Mann-Whitney U test indicated that there was a statistically significant difference, with $U(N_{\text{Sweden}} = 139, N_{\text{United States}} = 32) = 1512.5, z = -2.919, p = 0.004$.

The importance of reporting when noticing something unusual with the car was compared between Sweden and the United States. On average, Sweden (Mdn = 4) was higher than the United States (Mdn = 3). A Mann-Whitney U test indicated that there was a statistically significant difference, with $U(N_{\text{Sweden}} = 139, N_{\text{United States}} = 32) = 1716, z = -2.156, p = 0.031$.

The importance of getting an explanation when a function isn't working as expected was compared between Sweden and the United States. On average, Sweden (Mdn = 4) was higher than the United States (Mdn = 3.5). A Mann-Whitney U test indicated that there was a statistically significant difference, with $U(N_{\text{Sweden}} = 139, N_{\text{United States}} = 32) = 1726, z = -2.154, p = 0.031$.

The importance of being able to share an idea about a car function or feature was compared between Sweden and the United States. On average, Sweden (Mdn = 3) was equal to the United States (Mdn = 3). However, a Mann-Whitney U test indicated that there was a statistically significant difference-, with $U(N_{\text{Sweden}} = 139, N_{\text{United States}} = 32) = 1722.5, z = -2.082, p = 0.037$.

Figures 7.16, 7.17, 7.18 and 7.19 illustrate the distribution of how the participants of the two nations answered the questions that revealed a significant difference presented above. The mean rank is calculated by first collating and ranking all the data from both groups. Each response is then assigned a rank. To determine the mean rank for each group individually, the sum of the ranks are calculated and divided by the number of responses in the group (Sum of ranks/ (N_{Group} = mean rank). The mean rank depends on how the scores are spread out and how often each score appears in the collated dataset. The sample sizes were 139 for Sweden and 32 for the United States. The mean rank in the figures provides a comparison of the responses between the two independent groups, offering an understanding into regional differences of their responses. A higher mean rank indicates that the group's responses generally have higher values on the question scale compared to the other group. Participants rated each question on a Likert scale from 1 to 5.

Figure 7.16 shows a higher mean rank for the United States at 108.23 compared to Sweden's mean rank at 80.88. In Figure 7.17 Sweden has a higher mean rank at 89.65 while the United States has a mean rank of 70.13. Figure 7.18 shows a higher mean rank for Sweden at 89.58 compared to the United States' mean rank of 70.44. Lastly, in Figure 7.19 Sweden has a higher mean rank at 89.61 while the United States has a mean rank of 70.33. See Appendix B for the full questions labeled on the y-axes in the figures below.

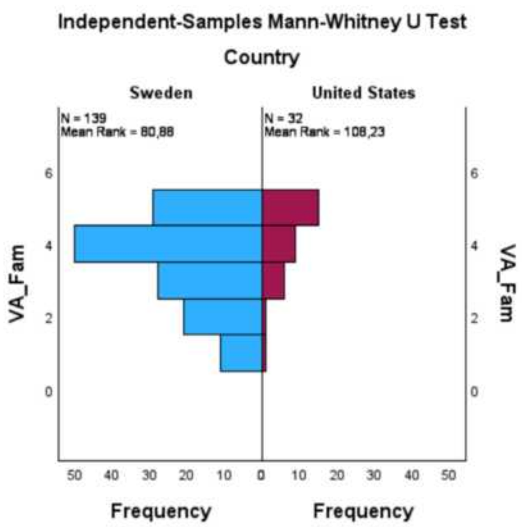


Figure 7.16: Participants familiarity of voice assistants, comparing Sweden & US

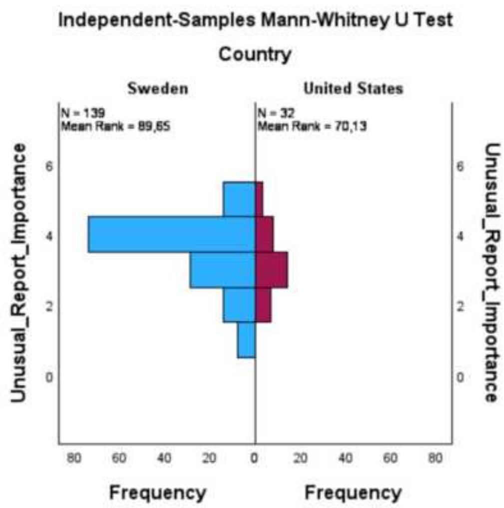


Figure 7.17: Participants perceived importance of reporting in an unclear situation, comparing Sweden & US

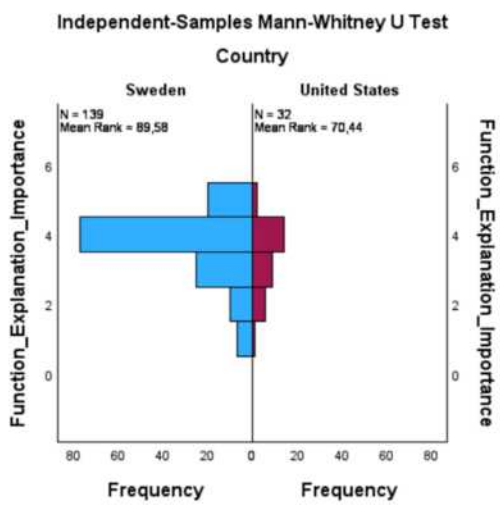


Figure 7.18: Participants perceived importance of receiving an explanation in a confusing situation, comparing Sweden & US

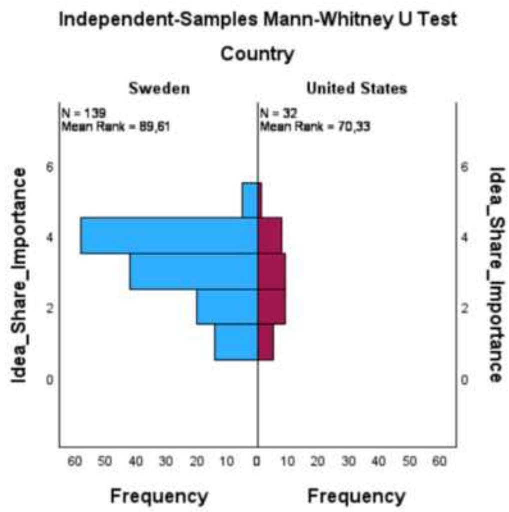


Figure 7.19: Participants perceived importance of sharing an idea, comparing Sweden & US

7.1.4.2 Relations between questions

The correlation matrix, figure 7.20, visualize a comprehensive representation of the relationships between the survey questions. Each row and column in the matrix corresponds to a question in the survey, identified with the same ID number. Each cell in the matrix represents the correlation between two questions. Spearman's rank correlation ranges from 1 to -1 where the coefficient value near 1 is a very strong positive correlation and a coefficient value near -1 is a very strong negative correlation. A correlation coefficient of 0 indicates on no correlation, see the color coding levels in Figure 7.20. Cells marked with a double asterisk (**) indicate a highly significant correlation with a very low probability (< 0.01) being due to chance. Cells marked with a single asterisk (*) indicate a significant correlation with a low probability (< 0.05) being due to chance.

The results show that the Correlation matrix seems concentrated around the diagonal. This indicates correlations between the key solutions(explanation, report, and solve) within the scenario questions and not across them.

Spearman's rank correlation reveled three cases of very strong positive correlations where the coefficient values are ≥ 0.8 and 24 cases of strong positive correlations with a coefficient value at ≥ 0.7. There are 41 cases of negative coefficient values, however, these values are close to 0 and are considered as no correlation.

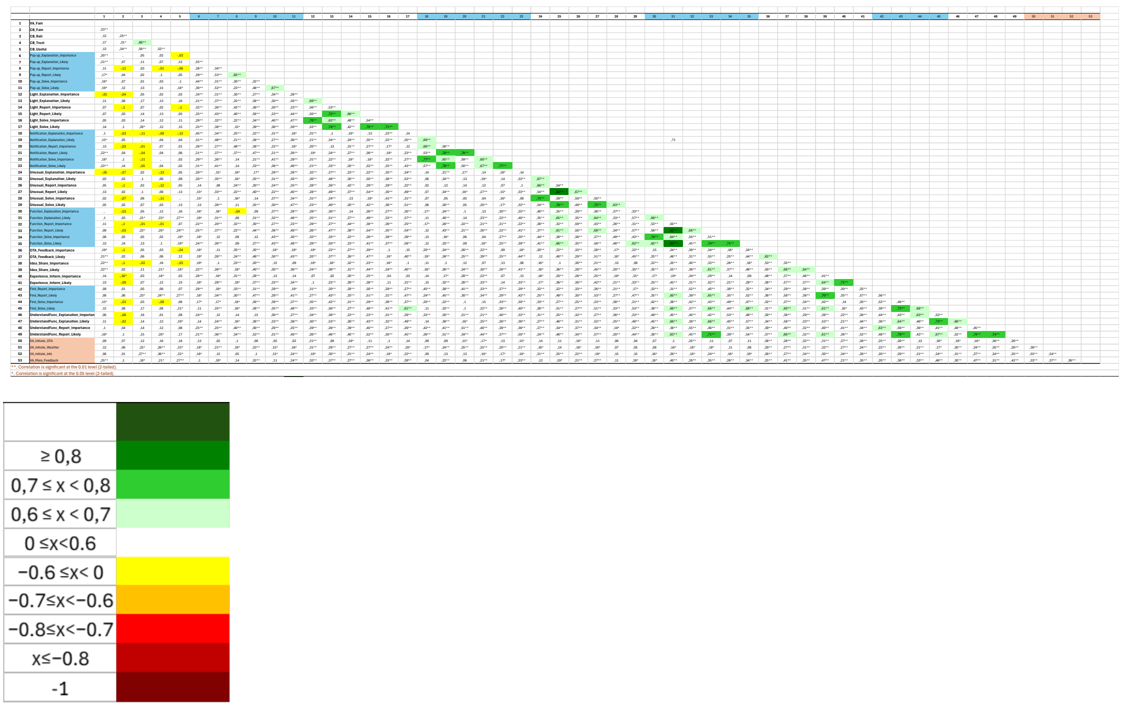


Figure 7.20: Correlation matrix, illustrating relations between survey questions (zoom in to view detailed relations)

Figure 7.21 and 7.22 are zoomed-in sections derived from the larger correlation matrix, Figure 7.20, highlighting the cases with coefficient values at ≥ 0.8. Figure 7.21 shows a strong positive correlation of 0.8 between the likeliness of using a VA in the statements: *"When I notice something unusual with my car, I would like an* ***explanation*** *of what the problem is"* and *"When I notice something unusual with my car, I would like to* ***report*** *this problem to Volvo"*. The correlation indicates on a highly significant correlation with a very low probability (< 1%) being due to chance.

		1	2	3	4	5	6
1	Unusual_Explanation_Importance						
2	Unusual_Explanation_Likely	,67**					
3	Unusual_Report_Importance	,66**	,54**				
4	Unusual_Report_Likely	,54**	,80**	,67**			
5	Unusual_Solve_Importance	,73**	,59**	,59**	,50**		
6	Unusual_Solve_Likely	,54**	,78**	,49**	,75**	,63**	

Figure 7.21: Correlations in situations of noticing something unusual with the car

Figure 7.22 highlights two strong positive correlations.The first, with a correlation of 0.82, is between the likeliness of using a VA in the statements: *"When a function isn't working as expected, I would like an* ***explanation*** *of what the problem is"* and *"When a function isn't working as expected, I would like to* ***report*** *this problem to Volvo"*. The second, with a coefficient value of 0.83, is between the likeliness of using a VA in the statements" *"When a function isn't working as expected, I would like an* ***explanation*** *of what the problem is"* and *"When a function isn't working as expected, I would like to know how to* ***solve*** *it (such as help guidance troubleshooting)"*. These correlations indicate a highly significant correlation with a very low probability (< 1%) being due to chance.

		1	2	3	4	5	6
1	Function_Explanation_Importance						
2	Function_Explanation_Likely	,69**					
3	Function_Report_Importance	,53**	,50**				
4	Function_Report_Likely	,56**	,82**	,68**			
5	Function_Solve_Importance	,74**	,64**	,50**	,51**		
6	Function_Solve_Likely	,60**	,83**	,45**	,74**	,71**	

Figure 7.22: Correlations in situations when a function is not working as expected

7.2 Qualitative data finding

All participants, regardless of their previous responses, were given the opportunity to answer to open-ended question regarding voice assistants about their likes and dislikes. Among those who responded two primary advantages were highlighted. First, many mentioned the appreciation of hands- and vision-free interface, and particularly highlighting its contribution for increased safety in vehicles. Second, many responded with the advantage of ease of use of such application. In contrast, three significant disadvantages were identified. The responses showed that the most common issue was the voice assistant often misunderstanding the user. Other common dislikes included a high error rate and inaccuracies, along with frustrations about limited capabilities or functionalities not working as expected.

7.2.1 Findings from Thematic Analysis

The thematic analysis includes all qualitative responses, covering both the open-ended questions from the survey and the responses from the expert interviews. To distinguish between the two, all survey responses are displayed on green post-it notes, while the survey responses from the expert interviews are shown on pink post-it notes in Figures 7.23, 7.24, 7.25, 7.26, 7.27 and 7.28. Each figure is divided into a category, has an overarching theme, a descriptive text and five or six post-it notes from the dataset.

Figure 7.23 shows opportunities with VAs that could be made through the advancements in generative AI. This question was only asked to the experts and therefore there are only pink post-it notes in this theme.

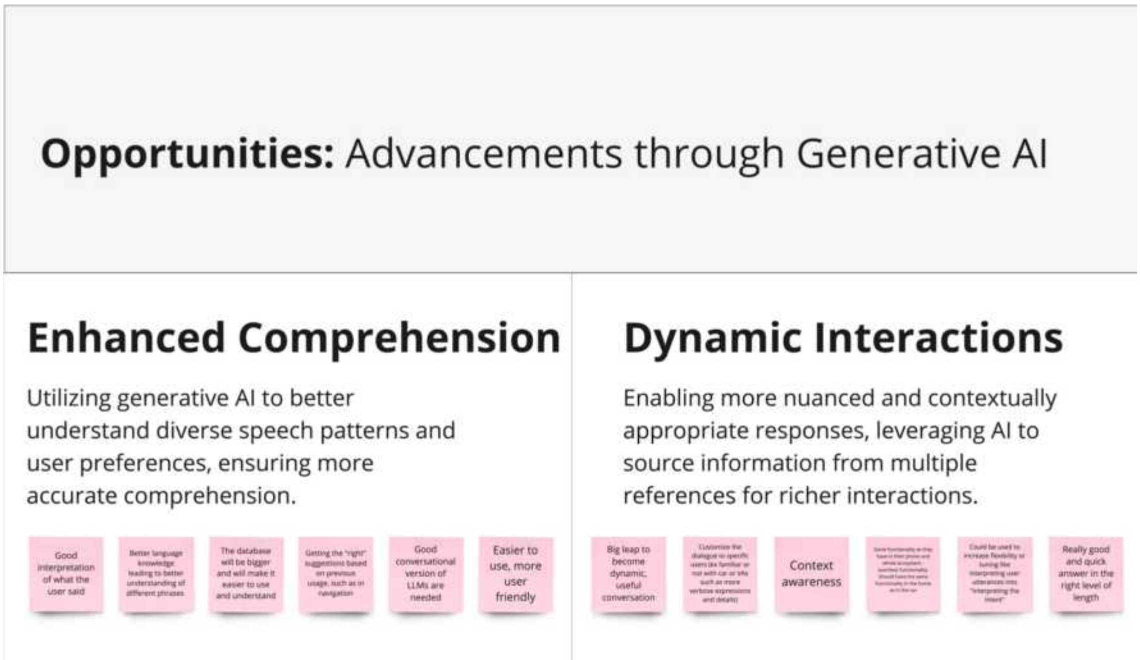


Figure 7.23: Identified opportunities through advancements in generative AI

Figure 7.24 demonstrate the capabilities of current VAs. The focus here is on User Experience and Accessibility, capturing aspects of VA that respondents appreciated. As previously mentioned, the green post-it notes display the survey responses and the pink ones display expert interview responses.

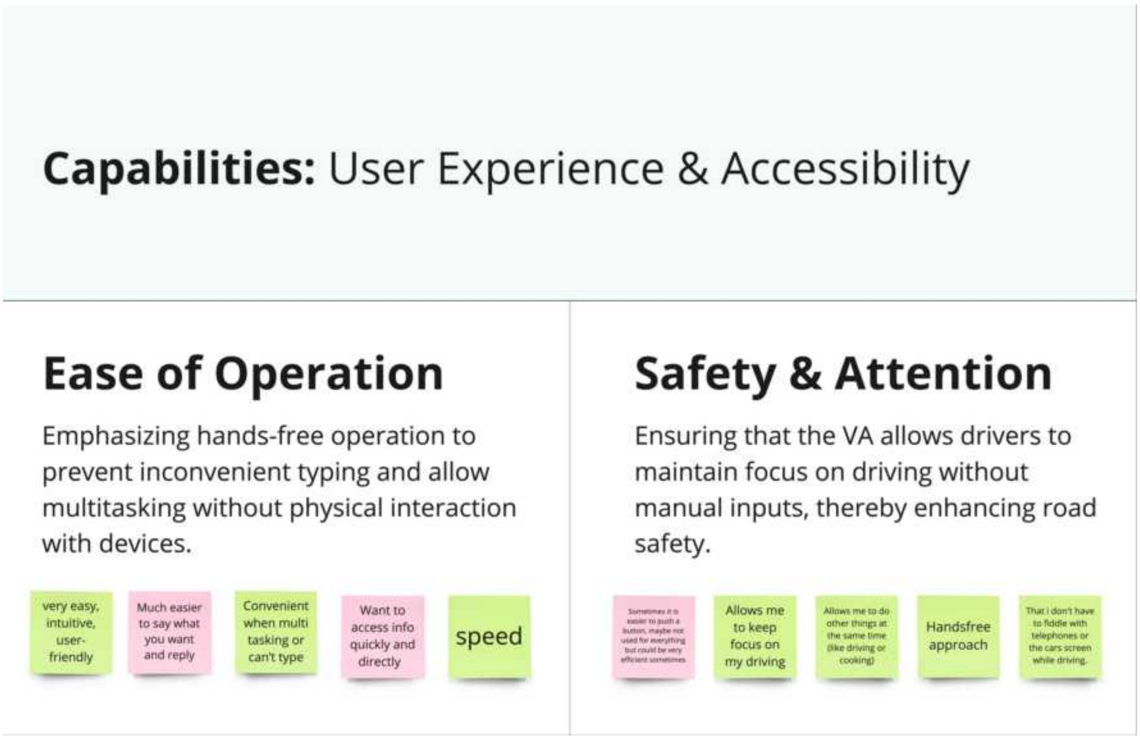


Figure 7.24: Identified capabilities of current VAs

Figures 7.25 and 7.26 present the user needs and desired functionality in VAs. The user needs cover Vehicle Integration and functionality as well as contextual and environmental awareness.

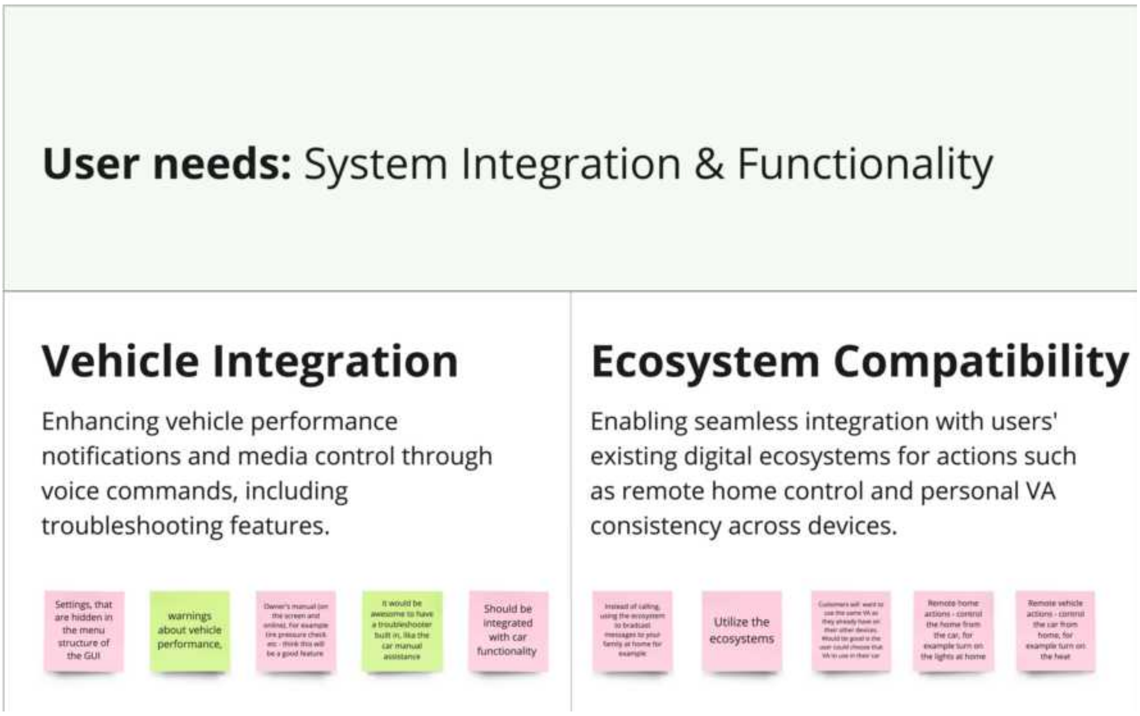


Figure 7.25: Identified user needs: System integration & functionality

Figure 7.26: Identified user needs: Contextual & environmental awareness

Figure 7.27: Identified privacy & data security challenges

Figure 7.27 cover the challenges associated with AI-powered VAs in vehicles. This includes the privacy and data security aspect. Figure 7.28 showcases the limitations in the current state of VAs, these are presented as the performance and reliability aspect.

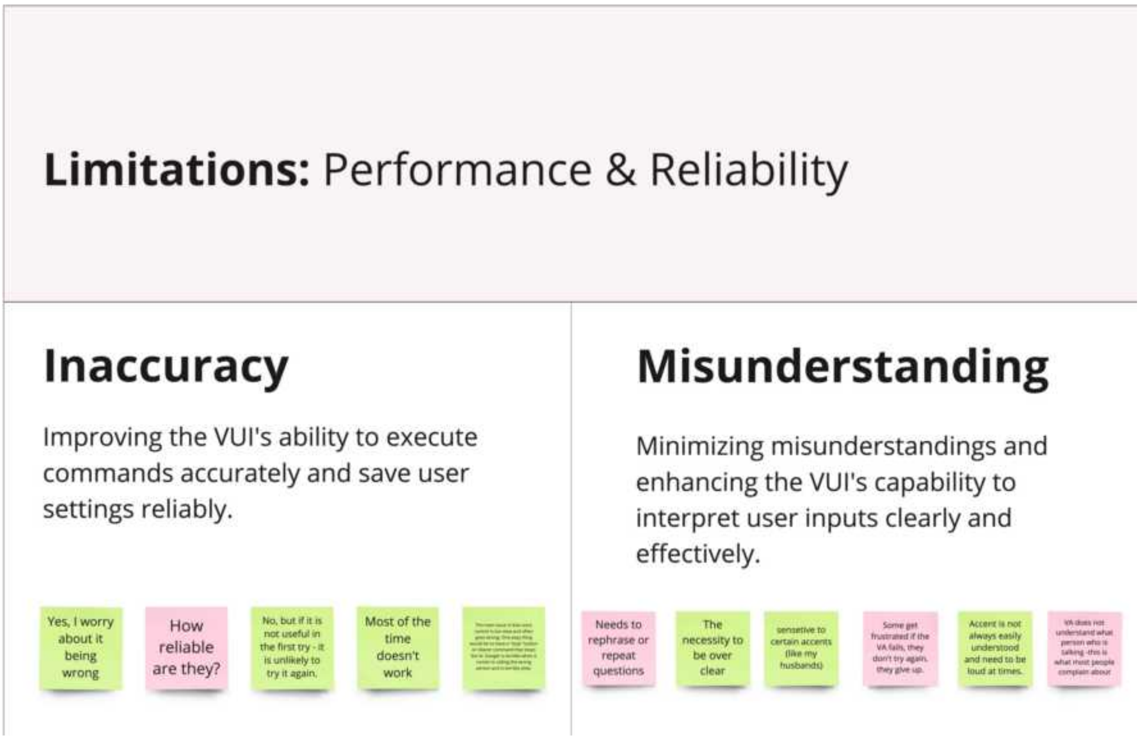


Figure 7.28: Identified performance & reliability limitations

8

Discussion

This chapter discusses the implications of the study findings and explores potential pathways for further enhancement of VAs. There will also be a critical and evaluative discussion regarding the methodological approaches of the project and their impact on the end result. Lastly, this chapter proposes directions for future research.

8.1 Current Capabilities & limitations of VAs

The study is initiated by exploring the participants' relationship with VAs, examining both their capabilities and limitations. This exploration aimed at gaining a deep understanding of the current experiences with VAs. By assessing how participants use and feel about these technologies, the study tried to identify areas where voice assistants succeed in meeting user expectations and where they fall short. These findings lay the groundwork for answering the research question: *Is there a need for AI-enabled in-car voice user interfaces, and what specific functionalities are drivers seeking?* By establishing what users valued in current VAs as well as their limitations, insights were gained into whether there is a user need for more advanced VAs.

8.1.1 Capabilities

The results of our interviews and surveys clearly indicated that user experience, defined by ease of operation and safety, is highly valued in the acceptance and effectiveness of VAs in vehicles.

Participants frequently highlighted ease of operation provided by VAs as one of the most important features, as shown in Figure 7.24. They especially point out hands-free operation to prevent inconvenient typing and allowing drivers to perform tasks without physical interaction with devices. This capability is not only convenient but crucial for allowing multitasking in a way that does not compromise the driver's attention or safety. The findings highlight the importance of VUI to prioritize minimalistic interaction designs that reduce cognitive load on users.

Participants of the study also highlighted that voice interfaces can enhance road safety, as shown in Figure 7.24. It was found that the ability of voice interfaces to allow drivers to maintain focus on driving without manual inputs is a key advantage. This aligns with previous research on VUIs in vehicles where it is stated that VAs can reduce distractions and improve safety in traditional driving as it lowers the cognitive load by handling non-driving tasks [26]. Therefore, the integration of VAs in vehicles should not only focus on functionality but also on how these systems can maintain or improve a driver's situational awareness. While multitasking could be seen as a natural part of modern driving, our findings suggest that VUIs can make this multitasking safer. Nevertheless, it is important to consider the cognitive load related to using VAs. Current voice interfaces, while allowing for multitasking, need to be carefully designed to make sure they do not mistakenly increase cognitive load by requiring complex voice commands or delivering too long or complicated responses.

8.1.2 Limitations

The feedback from participants regarding the limitations of VAs accentuate challenges in performance and reliability. These challenges are derived from issues with misunderstanding and inaccurate responses. Each of these issues undermines the effectiveness of VAs.

Participants frequently mentioned misunderstanding by VAs as one of the current major issues, shown in Figure 7.28. A notable difficulty in VA accurately processing accents, recognizing spoken words, and interpreting various phrasing styles. This issue both frustrates users as well as reduce the likeliness of continued use of the technology. The severity of these misunderstandings points to a need to enhance language models within VAs. Such improvements can include diversifying training data to encompass a variety of dialects and speech patterns, aiming to accommodate the global variability in language usage more effectively.

It was also found that inaccurate responses is a critical concern of users, shown in Figure 7.28. Errors such as misinterpreting addresses for navigation or dialing the wrong contacts likely cause inconvenience and is also a potential safety risk, especially in driving contexts. This issue accentuate the need for VAs to achieve higher accuracy in voice recognition and context understanding. Future improvements could include more advanced and context-aware systems that can interpret the intent behind user commands more reliably, thereby reducing errors in their responses.

8.1.3 Summary

To summarize the capabilities and limitations of VAs in vehicles it is pointed out that most participants thinks VAs are useful when they work as expected. However, the main challenge is that these assistants often fall short of expectations. The key issues include frequent misunderstandings of user commands and the inaccurate responses. This gap between expectation and performance indicate the need for improvements in the reliability and accuracy of VAs to improve user experience in vehicles. Thereby, when they work as expected, users find VAs to be a very convenient way of carrying out tasks and receiving information while driving.

8.2 User needs

The result from the survey responses and the expert interviews provided a comprehensive understanding of user needs and what specific functionalities they are seeking in an in-car VA. Overall, the study revealed that 56%, as shown in Figure 7.15, of the participants had a positive attitude using a in-car VA and that they would be more engaged in providing feedback, asking questions, or report problems if they could do so using a VA. The following sections will present and discuss these key findings about user needs and desired features. These findings are strongly related to the research question: *Is there a need for AI-enabled in-car voice user interfaces, and what specific functionalities are drivers seeking?* In the following section, both user needs and functionalities are discussed. It also includes a discussion on regional differences in subsection 8.2.1.2, providing an answer to the research question *Are there any differences in attitudes towards using in-car voice assistants between individuals in the United States and Sweden?*.

8.2.1 System Integration & Functionality

A significant finding, presented in Figure 7.25, showed that users would like a system capable of providing vehicle performance notifications, troubleshooting features and control media through voice commands. The results in section 7.1.1 shows the user needs for utilizing a VA in different scenarios. It indicates a user demand for support in situations involving feature and functions confusion. This includes seeking support through guidance when having trouble finding how to perform a task in the car. Additionally, users expressed a need for guidance or obtaining explanations when a function is not working as expected. The participants also found it critical to get support from a VA when noticing something unusual in the car, such as such as odd noises, vibrations, or changes in the car's handling. In these scenarios participants indicated both a high importance and a high likeliness of using a Car Co-pilot for getting assistance through explanations and troubleshooting, see Figures 7.6 and 7.8.

The most significant result from the survey highlighted the importance of warnings related to performance and high-severity problems. In Figures 7.6, 7.7 and 7.8 question 2 (Q2:'When I see a warning light showing in the dashboard of my car') received the highest rated on the importance of addressing the problem and the likeliness of using a Car Co-pilot. The responses of this scenario stands out compared to the other scenarios, indicating on a strong user need for support through a VA in high-severity situations.

Furthermore, three very strong correlations were found between the solutions (Explain, Report & Solve) in two scenarios regarding the likeliness of using of a Car Co-pilot. The correlations were found when a function is not working as expected (Q5) and when the driver does not understand a problem with the car (Q4). These correlations reinforces the user needs of using a Car Co-pilot in confusing and vague situations. Moreover, as presented in Figures 7.21, 7.22 a pattern of correlations was highlighted between the key solutions specific to individual scenarios, rather than across different scenarios. This outcome is likely due to the structure of the survey, which included a variety of scenarios that were not closely related. However, some weaker correlations were found **across** scenario questions related to feature and function confusion (Q5, Q6 and Q7). Upon reviewing these correlations, it is clear that these questions share similar characteristics. The results indicate that the users perceived resemblance in these scenarios and therefore provided similar responses to the questions.

Finally, to address some of these issues, participants suggested a system with integrated car functionality and owner's manual. Integrating functionality that holds information about the car and have the ability to troubleshoot problems could effectively address problems that occur in the car. It could also reduce unnecessary visits to repair service shops and enhance UX.

Providing user feedback, such as providing feedback after a recent OTA update, sharing ideas or inform about an experience, was generally rated lower in importance and likeliness of using a Car Co-pilot. Although, 40 percent still considering it very- or extremely important and likely. This suggests that the results remains significant even though it not the highest priority. However, compared to the other scenarios, this indicates that the users are more focused on situations were they encounter problems and need assistance, favoring their driver experience over actions that benefit the manufacturer. According to Liu [24] and Balan [18] understanding the driver and their preferences provides valuable insights and improve customer satisfaction. Therefore, encouraging and promoting user feedback could be a effective strategy to achieve this [24].

Lastly, the participants were moderately positive, see Figure 7.14, towards having a Car Co-pilot initiating the conversation. However, they answered that they would not enjoy using the feature if it became a disturbance.

8.2.1.1 Ecosystem compatibility

Participants, particularly those from the expert interviews shown in Figure 7.25, highlighted the importance of seamless integration of digital ecosystems across devices. This includes utilizing their existing VA regardless of location and device. This involves remote home actions, such as controlling the home from the car i.e, turning on the lights at home or sending messages to a home assistant for family members. It also includes remote vehicle actions, such as controlling the car from home i.e, preheating it before use. Experts also mention that users may already be familiar with one type of VA ecosystem, such as Google Assistant, Alexa, or Siri. This familiarity can benefit users by making it easier for them to navigate and utilize these technologies. As users become accustomed to specific commands and interactions within one system, they can transfer some of that knowledge to other, similar systems, improving their overall user experience and efficiency. Moreover, this prior experience can reduce the learning curve associated with new devices and services within the same ecosystem, leading to quicker adoption and more profound integration into daily routines.

8.2.1.2 Regional differences

Some differences in the responses between the United States and Sweden were found to the questions asked in the survey. The Figures 7.10, 7.11 and 7.12 in Chapter 7 showed that responses from the United States overall perceived the problem situations in the scenarios as less important and that the use of a Car Co-pilot in these situations was less likely compared to the responses from Sweden. However, both countries individually ranked the highest on the high-severity scenario, Q2 'When I see a warning light showing in the dashboard of my'. This indicates that there is a user need regardless of regional difference.

In four cases shown in section 7.1.4.1 there were significant differences between the responses of the United States and Sweden. A significant difference was captured in the familiarity of using VAs where the United States' responses indicated a higher familiarity than the Sweden's responses. However, these high rated responses did not reflect a higher likeliness of using a Car Co-pilot in the scenarios compared to the Swedish responses. Moreover, the Swedish responses indicated a higher importance of reporting when discovering something unusual with the car, as well as higher importance of getting an explanation of what the problem is when a function is not working as expected. The Swedish responses also indicated a higher importance of being able to share a feature or function idea to Volvo. These findings indicate that the Swedish drivers perceives these three case situations as more important than the United States.

These findings may indicate that there is differences in the market and that services might need to be adapted accordingly. However, since we stopped our research at this point, and further research would be necessary to understand underlying reasons to these differences.

8.2.2 Contextual & environmental awareness

Participants from the survey and expert interviews expressed their desire for a in-car VA that possesses contextual and environmental awareness, as shown in Figure 7.26. This refers to the car's ability to sense, interpret and reasons to its surroundings in various situations. Contextual awareness involves the VA understanding the driver's habits and preferences, such as learning about preferred routes or specific settings within the car. Environmental awareness covers the car's ability to interact with its external environment, such as providing real-time traffic updates and suggest alternative routes as needed. Systems like these would not only enhance safety by minimizing the distraction of a screen, but also improve user convenience and comfort by making the driving experience more personalized and adaptive.

8.3 Opportunities

The implementation of Generative AI within voice assistants comes with various opportunities to overcome current limitations and enhance user experience. This section discusses the implications of these opportunities, particularly focusing on how advancements in Generative AI can contribute to enhanced comprehension and dynamic interactions. These insights address the research question: *Does Generative AI have any potential in improving in-car Voice Assistants?*

8.3.1 Enhanced comprehension

Enhanced comprehension means using generative AI to help understand diverse speech patterns and user preferences, to provide more authentic understanding. This is possible through the use of LLMs, a technology within generative AI, as they more effectively can interpret new or different words by understanding the context in which they are mentioned. This is especially central in driving contexts where clarity and accuracy are crucial for safety and usability. For example, enhanced comprehension can reduce errors in conversation during demanding driving situations, such as navigating in stressful traffic situations or operating vehicle controls through voice. The language understanding is also more nuanced due the diverse and large amount of training data in the dataset [31], [32]. This helps VAs respond more accurately to questions, recognize the different phrasing, and keep track of context in longer conversations. However, even in cases with limited data, LLMs can enhance the comprehension of VAs. The enhancement is possible due to using tokens as a model or detailed map of languages, allowing it to navigate and grasp human communication more efficiently.

Furthermore, the advancement in generative AI can lead to great improvements of the overall experience of VAs. Experts mention, as shown in Figure 7.23, that these advancements made within generative AI and LLMs will make the VAs easier to use and more user friendly. They also mention that leveraging LLMs can help VAs better cater to different user preferences and speech patterns. The improved comprehension enables a further personalized experience and could reduce the learning curve for users [22]. Over time, the VA can adapt to the user's way of communicating, favorite topics or understand subtle preferences based on previous interactions. Thereby, VAs can offer more tailored responses and anticipate user needs more effectively. For instance, a VA that remembers a user's preferred vehicle temperature or seating arrangements can automatically be adjusted for a more personalized driving experience.

8.3.2 Dynamic interactions

Dynamic interactions focuses on facilitating more nuanced and contextually appropriate responses, using AI to retrieve information from multiple sources for richer interactions. Experts mention that VAs utilizing LLMs could increase flexibility by better interpreting user utterances and accurately determining the user's intent. (Figure 7.23). The comprehension of intent is achievable through the LLMs' ability to measure the importance of tokens [32]. Experts claim that intent recognition is central in improving the user experience. By understanding the users intent, the VA can provide more relevant and personalized responses, which facilitates more satisfying and engaging user experience. This is particularly important in a car driving context, where the ability to respond to the users needs in a timely and appropriate manner can significantly impact customer satisfaction. This is especially important in driving scenarios where the timing and relevance of a response, whether concise during stressful conditions or more detailed at a standstill, can significantly impact driver satisfaction and safety.

Moreover, as LLMs can combine multiple sources to produce and give more accurate responses using technologies as RAG [35]. This is especially useful in situations where the users' asks complex questions that require up-to date information or using multiple sources to find a truthful answer. The ability of LLMs to combine data from various inputs into responses that are consistent and suitable to the context further demonstrates their potential in enhancing user experience. To further engage users, experts mention the use of advanced natural language processing techniques to analyze, comprehend, and produce human-like text. This involves understanding grammar, syntax, and semantics, which are essential for VAs to produce language that feels natural and interesting to users.

The ongoing development of Generative AI and its application in vehicle VA promise to transform the interaction between drivers and their vehicles. As these technologies continue to advance, they will not only address current limitations but also open new avenues for creating more intelligent, responsive, and user-centric automotive systems.

8.4 Challenges

While the use of generative AI and LLMs presents great opportunities to improve user experiences, they also introduce challenges that need to be addressed, as shown in Figure 7.27. The experts specifically mention the phenomenon of 'hallucination' in LLMs as one of the main threats to implement this new AI technology in VAs. Hallucination occurs when models produce irrelevant or inaccurate information when trying to create a coherent and well-articulated response [31]. The inconsistency where the model confidently produces high quality text with inaccurate information, convincing the user that the provided information is accurate can harm the reliability of the VA. Which in turn might also have a negative impact on the manufacturers' reputation.

Another challenge mentioned by both survey participants and experts concerns privacy and data handling, as shown in Figure 7.27. Specifically empathizing user concerns in data privacy issues, such as unauthorized listening and data storage practices. Participants express concerns about personal integrity violations, especially when devices are always active, with microphones in their homes or cars, coupled with unclear control over where their data finally is stored. Another concern regards potential cyber attacks. Experts discuss where the data is sent in the case of cyber attack data breach, emphasizing the need for strong security measures. Additionally, there are broader worries about the ownership and control of AI development. As companies, governments, and others might take control, the risk of "wrong" persons gaining power without public awareness is a common worry. Such a scenario could potentially lead to privacy invasions if these people gain uncontrolled access to large amounts of user data. This complex landscape of concerns underscores the need for more transparent practices in AI development and data management.

8.5 Limitations

This section presents the limitations of the project, acknowledging the factors that may have affected the accuracy of the result.

8.5.1 Convenience sampling

The survey was conducted with employees at Volvo Cars who are a part of the CoDev program, which involves looking at vehicle enhancement and development. As participants in the program, they are more likely to have a higher level of technical knowledge than an average consumer. This could lead to overlooking problems that are more evident to regular users or to place higher importance on features that aligns with their professional experience or the manufacture's priorities. Additionally, the experts from the interviews were generally more familiar with using VAs and they showed a positive attitude towards them. This familiarity and positivity might have resulted in a biased and perhaps exaggerated beliefs in the importance of integrating VAs in cars.

Furthermore, the demographic distribution of the participants, presented in Section 7.1, could also have an impact on the outcome. Over 75% of the participants were men and approximately 85% were 45 years old or older. This demographic distribution may limit how well the result can be applied to women and could primarily reflect the experiences and behaviours of middle-aged men, excluding the perspectives and conditions of young adults.

The convenience sampling method was chosen in this study due to the ease of access of a larger group of participants. However, this approach might have led to biases, resulting in skewed responses. Therefore, the insights derived from this group need to be considered cautiously, as it can not be concluded that they reflect the market needs or consumer preferences at large.

8.5.2 Methodological implementation

The study did not fully implement the Triptech method guidelines, including conducting follow-up interviews with focus groups and storyboards to visualize interactions. The adaptation of the Triptech method was a conscious decision to better suit the automotive context, such as testing several solutions to a problem (Explain, Report, and Solve) and include an informative text on the technology that could mitigate the issue. However, the decision to not include follow-up interviews with survey participants in focus groups was made due to our limited time to setup, conduct and analysis. Instead, expert interviews were favored as a good trade-off to still delve deeply into users' needs and wants while keeping the preparation and analysis time manageable within our timeframe.

The study did not include real-world testing, as the primary goal of the Triptech method was to investigate whether there is a need for the technology at all. This initial study aimed to determine if AI-powered VAs are even desired by users. If they are not, there would be no need for empirical testing.

By excluding real-world testing, the study focused on assessing the potential demand for the technology through hypothetical scenarios. While this approach provides preliminary insights into user preferences and needs, it may limit the depth of the findings compared to what could be gathered from real-world interactions. Hypothetical scenarios can offer initial ideas, but they may not capture the nuances and complexities of actual user behavior. Therefore, the insights derived from this study might differ from those obtained through empirical testing in real-world settings. Despite this limitation, the study serves as a valuable first step in understanding the potential acceptance and demand for AI-powered VAs, laying the groundwork for future research that may include more extensive empirical testing.

8.6 Evaluation of the Triptech method

This section evaluates the usage of the Triptech method, gives recommendations for its future use, and discusses the contributions made to extend the Triptech method. It focuses on adaptations made during this study to enhance the effectiveness of the Triptech method in various contexts, particularly within the automotive industry.

8.6.1 Use & recommendation

Utilizing the Triptech method for a evaluating a design concept within the automotive context was challenging and needed substantial tailoring and adaptation to suit the context. There was also limited information on the different applications of the Triptech method which further obstructed the work. This lack of guidance made it difficult to determine whether our study stuck to the Triptech method or if it was modified to the extent that we developed our own guidelines.

For further evaluation of design concepts in the automotive context, we recommend to adapt the parameters to suit the study, as in this case, scenario, likeliness and importance. This approach proved effective in understanding the desirability of getting assistance from VAs in various situations. Replacing the parameter frequency/likeliness with the most suitable parameter for the specific study in question. Also, using explain, report and solve to evaluate the solutions for each scenario may need adaptation to suit research of other design concepts within the automotive environment.

To summarize, the Triptech method is a suitable method to capture early user needs and desires of design concepts. However, it may require tailoring of the frequency parameter to align with the study's goals. In this study, creating three possible solutions for each scenario was appropriate, but this approach can be adapted to investigate single or multiple solutions depending on the research context.

8.6.2 Contribution

This thesis has contributed to extending the Triptech method by adapting it to an automotive context and addressing the specific needs and challenges associated with in-car voice assistants. The traditional Triptech method focuses on evaluating early design concepts through triangulation, prototyping, and technology. This study has tailored the method to better suit the unique requirements of in-car environments.

One of the key contributions is the focus on testing several solutions rather than a single one. This approach investigates the complexity and variability of user needs in automotive settings. By presenting different solutions, such as explaining a problem, reporting it, or providing guidance to solve it, the study ensures that the method can capture a broader range of user preferences and needs. This provides more comprehensive insights into the desirability of in-car voice assistants, making the method more robust and adaptable to different contexts.

Another adaptation is the replacement of the frequency parameter with the likeliness of using an in-car voice assistant. In the automotive context, the frequency of car issues is relatively rare, but the importance and liklieness of seeking assistance can be high. By focusing on the likeliness of using a voice assistant in specific scenarios, the study provides a more relevant measure of user needs and preferences. This helps in understanding the problems users face and their willingness to use technological solutions to solve these problems.

In summary, this thesis has extended the Triptech method by introducing an approach with multiple solution testing, replacing the frequency parameter with the likeliness of use, and tailoring the method to the specific context of in-car voice assistants. These adaptations make the Triptech method more flexible, user-centered, and suitable for evaluating design concepts in the automotive industry .

8.7 Future work

The use of the convenience sampling method in this study might have limited the ability to fully capture the needs and perspectives of actual users of in-car VAs. Involving a more varied group of users in a future research would provide unique and valuable insights from their everyday experiences, reflecting the actual user needs. This approach could facilitate the development of more user-centred and useful concepts, which potentially could increase market success. Additionally, involving a broader and more diverse group of users, with varied age, gender, technological knowledge and broader geographical locations, could capture a wider range of use cases and desired interactions with VAs. Therefore, future studies should aim to involve users with various backgrounds to enhance the design of a user-friendly in-car VA.

Moreover, the study did not incorporate real-world user testing, which could have an impact on the insight derived from this study. For future research, real-world testing should be considered for several reasons. It would provide real-time feedback in actual driving conditions, offering direct reactions to the VA's performance and user experience. Additionally, real-world testing could assess the VA's reliability and safety across different driving situations, where conditions may change rapidly. These tests would verify whether the system can handle various conditions and tasks seamlessly as well as testing the relevance of the VA's functions to ensure that it meet user needs and expectations.

9 Conclusion

This thesis has explored the evolving role of AI-enabled voice assistants in the automotive context. More specifically it set out to investigate the user need for in-car voice assistants and looking at attitudinal differences between individuals in the United States and Sweden. It also aimed at exploring the potential of generative AI integrated with these voice assistants. To gain a comprehensive understanding of the area, the study utilized a mixed method approach inspired from the TripTech method. Through a survey and expert interviews, both quantitative and qualitative data was collected. The qualitative data was analyzed through a thematic analysis in six phases, while the quantitative analysis used various statistical methods. The findings reveal a complex landscape of user acceptance and expectations. It was found that more than half of the survey participants had a positive attitude towards using an in-car VA. The technology was especially favourable in high severity and confusing situations where the users may need an explanation or guidance to solve and issue. Responses from the United States generally perceived the problem scenarios as less important and the likeliness of using a VA in different situations as less likely compared to the responses from Sweden. Additionally, four cases indicated on statistically significant differences between the United States and Sweden.

It was also found that users see potential in voice interfaces in cars as they offer a safer and more convenient way of carrying out tasks and receiving information while driving. However, the qualitative data that the current experience of using VUIs indicate that they do not fulfill user expectations when it comes to language comprehension and accuracy. A commonly mentioned situation is where the VA misunderstand a name and then dials the wrong contact. Most users find VAs valuable when they work as expected. Nevertheless, the main challenge is that these assistants often fall short of expectations. Several user needs and desired functionalities were identified such as system integration, ecosystem comparability and contextual and environmental awareness. Creating a more flexible and personalized experience was strongly emphasised to improve the voice interface.

Introducing AI powered VAs has shown to present both opportunities and challenges. Opportunities found are related to utilizing the LLM technology to enhance the VAs comprehension and facilitating more dynamic interactions with users, solving the main limitations of the current state of voice assistants. However, regarding inaccurate responses and privacy issues are still prevalent. To conclude, our study shows generative AI has potential in improving in-car voice assistants, yet challenges concerning privacy and inaccuracy of the produced materials remain and needs to be thoroughly looked at before integrated in new vehicles.

Bibliography

[1] C. Murad and C. Munteanu, “Designing voice interfaces: Back to the (curriculum) basics,” in *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, 2020, pp. 1–12.

[2] A. de Barcelos Silva, M. M. Gomes, C. A. da Costa, *et al.*, “Intelligent personal assistants: A systematic literature review,” *Expert Systems with Applications*, vol. 147, p. 113 193, 2020.

[3] J. Wang, W. Ma, P. Sun, M. Zhang, and J.-Y. Nie, “Understanding user experience in large language model interactions,” *arXiv preprint arXiv:2401.08329*, 2024.

[4] F. Weng, P. Angkititrakul, E. E. Shriberg, L. Heck, S. Peters, and J. H. Hansen, “Conversational in-vehicle dialog systems: The past, present, and future,” *IEEE Signal Processing Magazine*, vol. 33, no. 6, pp. 49–60, 2016.

[5] H. Heuvel, J. Boudy, R. Comeyne, S. Euler, A. Moreno, and G. Richard, “The speechdat-car multilingual speech databases for in-car applications: Some first validation results,” 1999.

[6] J. H. Hansen, P. Angkititrakul, J. P. Plucienkowski, *et al.*, “" cu-move": Analysis & corpus development for interactive in-vehicle speech systems.,” in *INTERSPEECH*, 2001, pp. 2023–2026.

[7] E. Levin, S. S. Narayanan, R. Pieraccini, *et al.*, “The at&t-darpa communicator mixed-initiative spoken dialog system.,” in *INTERSPEECH*, 2000, pp. 122–125.

[8] Capgemini Research Institute. “Voice on the go: How can auto manufacturers provide a superior in-car voice experience?” (2019), [Online]. Available: `https://www.capgemini.com/wp-content/uploads/2019/11/Report-%E2%80%93-Voice-on-the-Go.pdf` (visited on 05/16/2024).

[9] C. Mark. “Do brits want their cars with digital voice assistants?” (Feb. 2024), [Online]. Available: `https://business.yougov.com/content/48530-do-brits-want-their-cars-with-digital-voice-assistants` (visited on 05/15/2024).

[10] Consumer Electronics Show. “About ces.” (n.d.), [Online]. Available: `https://www.ces.tech/about-ces.aspx` (visited on 03/15/2024).

[11] Volkswagen-newsroom. “World premiere at ces: Volkswagen integrates chatgpt into its vehicles.” (2024), [Online]. Available: `https://www.volkswagen-newsroom.com/en/press-releases/world-premiere-at-ces-volkswagen-integrates-chatgpt-into-its-vehicles-18048` (visited on 03/05/2024).

[12] Mercedes-Benz Group. "The hyper-personalised user experience." (2024), [Online]. Available: https://group.mercedes-benz.com/innovation/product-innovation/technology/ces-2024.html (visited on 03/07/2024).

[13] BMW Group. "Bmw at the consumer electronics show (ces) 2024." (2024), [Online]. Available: https://www.press.bmwgroup.com/usa/article/detail/T0438991EN_US/bmw-at-the-consumer-electronics-show-ces-2024?language=en_US (visited on 03/05/2024).

[14] K. Seaborn, N. P. Miyake, P. Pennefather, and M. Otake-Matsuura, "Voice in human–agent interaction: A survey," *ACM Computing Surveys (CSUR)*, vol. 54, no. 4, pp. 1–43, 2021.

[15] N. J. Nilsson, *The quest for artificial intelligence*. Cambridge University Press, 2009.

[16] R. Chrisley, "Embodied artificial intelligence," *Artificial intelligence*, vol. 149, no. 1, pp. 131–150, 2003.

[17] R. Langevin, R. J. Lordon, T. Avrahami, B. R. Cowan, T. Hirsch, and G. Hsieh, "Heuristic evaluation of conversational agents," in *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, 2021, pp. 1–15.

[18] C. Balan, "Chatbots and voice assistants: Digital transformers of the company–customer interfacea systematic review of the business research literature," *Journal of Theoretical and Applied Electronic Commerce Research*, vol. 18, no. 2, pp. 995–1019, 2023.

[19] L. Liao, G. H. Yang, and C. Shah, "Proactive conversational agents in the post-chatgpt world," in *Proceedings of the 46th International ACM SIGIR Conference on Research and Development in Information Retrieval*, 2023, pp. 3452–3455.

[20] A. Schmitt, N. Zierau, A. Janson, and J. M. Leimeister, "Voice as a contemporary frontier of interaction design," in *European Conference on Information Systems (ECIS).-Virtual*, 2021.

[21] A. Pyae and T. N. Joelsson, "Investigating the usability and user experiences of voice user interface: A case of google home smart speaker," in *Proceedings of the 20th international conference on human-computer interaction with mobile devices and services adjunct*, 2018, pp. 127–131.

[22] T. Doi and Y. Nishikawa, "Usability analysis of smart speakers from a learnability perspective for novel users," *Applied System Innovation*, vol. 7, no. 3, p. 36, 2024.

[23] P. K. Murali, M. Kaboli, and R. Dahiya, "Intelligent in-vehicle interaction technologies," *Advanced Intelligent Systems*, vol. 4, no. 2, p. 2 100 122, 2022.

[24] J. Liu, F. Wan, J. Zou, and J. Zhang, "Exploring factors affecting peoples willingness to use a voice-based in-car assistant in electric cars: An empirical study," *World Electric Vehicle Journal*, vol. 14, no. 3, p. 73, 2023.

[25] R. Ringfort-Felner, M. Laschke, S. Sadeghian, and M. Hassenzahl, "Kiro: A design fiction to explore social conversation with voice assistants," *Proceedings of the ACM on Human-Computer Interaction*, vol. 6, no. GROUP, pp. 1–21, 2022.

[26] Y. Du, J. Qin, S. Zhang, S. Cao, and J. Dou, "Voice user interface interaction design research based on user mental model in autonomous vehicle," in *Human-Computer Interaction. Interaction Technologies: 20th International Conference, HCI International 2018, Las Vegas, NV, USA, July 15–20, 2018, Proceedings, Part III 20*, Springer, 2018, pp. 117–132.

[27] H. Dubberly and P. Pangaro, "On modeling what is conversation, and how can we design for it?" *interactions*, vol. 16, no. 4, pp. 22–28, 2009.

[28] Google Assistant. "Conversation design." (2021), [Online]. Available: `https://developers.google.com/assistant/conversation-design/welcome` (visited on 03/14/2024).

[29] K.-B. Ooi, G. W.-H. Tan, M. Al-Emran, *et al.*, "The potential of generative artificial intelligence across disciplines: Perspectives and future directions," *Journal of Computer Information Systems*, pp. 1–32, 2023.

[30] L. Banh and G. Strobel, "Generative artificial intelligence," *Electronic Markets*, vol. 33, no. 1, p. 63, 2023.

[31] N. Karanikolas, E. Manga, N. Samaridi, E. Tousidou, and M. Vassilakopoulos, "Large language models versus natural language understanding and generation," in *Proceedings of the 27th Pan-Hellenic Conference on Progress in Computing and Informatics*, 2023, pp. 278–290.

[32] M. Shanahan, "Talking about large language models," *Communications of the ACM*, vol. 67, no. 2, pp. 68–79, 2024.

[33] H. Zhao, H. Chen, F. Yang, *et al.*, "Explainability for large language models: A survey," *ACM Transactions on Intelligent Systems and Technology*, vol. 15, no. 2, pp. 1–38, 2024.

[34] H. Kotek, R. Dockum, and D. Sun, "Gender bias and stereotypes in large language models," in *Proceedings of The ACM Collective Intelligence Conference*, 2023, pp. 12–24.

[35] Amazon Web Services. "What is rag?" (2024), [Online]. Available: `https://aws.amazon.com/what-is/retrieval-augmented-generation/` (visited on 03/04/2024).

[36] Y. Hua, X. Xi, Z. Jiang, *et al.*, "Dialog-to-actions: Building task-oriented dialogue system via action-level generation," in *Proceedings of the 46th International ACM SIGIR Conference on Research and Development in Information Retrieval*, 2023, pp. 3255–3259.

[37] M. Gao, X. Hu, J. Ruan, X. Pu, and X. Wan, "Llm-based nlg evaluation: Current status and challenges," *arXiv preprint arXiv:2402.01383*, 2024.

[38] Persado. "Natural language generation: Use cases and business impact." (2022), [Online]. Available: `https://www.persado.com/articles/natural-language-generation/` (visited on 05/21/2024).

[39] K. Foster. "What is automatic speech recognition? a comprehensive overview of asr technology." (Sep. 2023), [Online]. Available: `https://www.assemblyai.com/blog/what-is-asr/` (visited on 03/29/2024).

[40] M. Malik, M. K. Malik, K. Mehmood, and I. Makhdoom, "Automatic speech recognition: A survey," *Multimedia Tools and Applications*, vol. 80, pp. 9411–9457, 2021.

[41] Y. Kumar, A. Koul, and C. Singh, "A deep learning approaches in text-to-speech system: A systematic review and recent research perspective," *Multimedia Tools and Applications*, vol. 82, no. 10, pp. 15 171–15 197, 2023.

[42] M. ajszczak, G. Cámbara, Y. Li, *et al.*, "Base tts: Lessons from building a billion-parameter text-to-speech model on 100k hours of data," *arXiv preprint arXiv:2402.08093*, 2024.

[43] I. Chen. "Why offline ai voice assistant is the future." (Oct. 2023), [Online]. Available: `https://mychen76.medium.com/why-offline-ai-voice-assistant-is-the-future-1d071b1b6422` (visited on 05/20/2024).

[44] iTnews Asia Team. "Toyota to use google cloud for next-gen offline voice assistants." (Oct. 2022), [Online]. Available: `https://www.itnews.asia/news/toyota-to-use-google-cloud-for-next-gen-offline-voice-assistants-586638` (visited on 05/20/2024).

[45] J. S. Ernst, S. Marton, J. Brinkmann, *et al.*, "Bias mitigation for large language models using adversarial learning," 2023.

[46] B. Davidson and R. Patel, "Forskningsmetodikens grunder," *Lund: Studentlitteratur*, pp. 28–31, 2003.

[47] R. Heale and A. Twycross, *What is a case study?* 2018.

[48] J. Bell and S. Waters, *Introduktion till forskningsmetodik*. Studentlitteratur, 2020.

[49] J. A. Séguin, A. Scharff, and K. Pedersen, "Triptech: A method for evaluating early design concepts," in *Extended abstracts of the 2019 CHI conference on human factors in computing systems*, 2019, pp. 1–8.

[50] P. Bahia. "An introduction to the ux triptech method." Accessed: 2023-05-26. (May 2023), [Online]. Available: `https://medium.com/@pazabahia/an-introduction-to-the-ux-triptech-method-7362fa02029f`.

[51] J. A. Séguin, *Triptech: A method for evaluating early design concepts*, Video file, Available at YouTube, Jul. 2019. [Online]. Available: `https://www.youtube.com/watch?v=05Gx1CFp0-Y`.

[52] S. McCombes. "Survey research | definition, examples & methods." (2023), [Online]. Available: `https://www.scribbr.com/survey-research-definition-examples-methods/` (visited on 03/17/2024).

[53] Statista. "Number of voice assistant users in the u.s. 2022-2026." (2023), [Online]. Available: `https://www.statista.com/statistics/1299985/voice-assistant-users-us/` (visited on 06/02/2024).

[54] A. Bogner and W. Menz, "The theory-generating expert interview: Epistemological interest, forms of knowledge, interaction," in *Interviewing experts*, Springer, 2009, pp. 43–80.

[55] H. Sharp, Y. Rogers, and J. Preece, *Interaction Design: Beyond Human-Computer Interaction*, 5th. Wiley, 2019.

[56] D. Stockemer, G. Stockemer, and J. Glaeser, *Quantitative methods for the social sciences*. Springer, 2019, vol. 50.

[57] M.-T. Puth, M. Neuhäuser, and G. D. Ruxton, "Effective use of spearman's and kendall's correlation coefficients for association between two measured traits," *Animal Behaviour*, vol. 102, pp. 77–84, 2015.

[58] T. W. MacFarland, J. M. Yates, *et al.*, *Introduction to nonparametric statistics for the biological sciences using R.* Springer, 2016.
[59] V. Braun and V. Clarke, *Thematic analysis.* American Psychological Association, 2012.
[60] IBM, *Ibm spss statistics algorithms*, `https://example.com/IBM_SPSS_Statistics_Algorithms.pdf`, Accessed: 2024-03-14, 2022.

A

Survey

This appendix presents the full survey, conducted in Microsoft Forms, that was sent out to CoDev drivers in the United States and Sweden.

AI-enabled voice assistants in vehicles

We're interested in understanding how voice-controlled technology can enhance your driving experience, focusing on safety, convenience, and vehicle management.

Your feedback is essential in identifying what features are most valuable and any concerns you may have. This information will guide the development of voice assistant capabilities that meet real-world needs.

Thank you for taking the time to fill this survey!

Team CoDev

* Required

About you

1

What is your gender? *

- Woman
- Man
- Non-binary
- Other
- Prefer not to say

2

How old are you? *

- 18-24
- 25-34
- 35-44
- 45-55
- 55>
- Prefer not to say

3

Where do you currently live? *

- ◯ Sweden
- ◯ United States

4

How familiar are you with using Voice Assistants (such as Google Assistant, Amazon Alexa, Siri)? *

- ◯ Never used it
- ◯ Use it a few times a year
- ◯ Use it a few times a month
- ◯ Use it a few times a week
- ◯ Use it daily

5

When you use Voice Assistants, what do you use them for? *

- ☐ **Hands-Free Communication:** Making calls or sending messages without using your hands; having messages read aloud to you and responding through voice.
- ☐ **Weather and News Updates:** Receiving real-time weather forecasts and current news briefings through voice commands.
- ☐ **Reminders and Alarms:** Using voice commands to set, manage, and receive reminders and alarms for your daily tasks and appointments.
- ☐ **Voice-Controlled Media:** Playing music, podcasts, and other media simply by speaking to your device.
- ☐ **Smart Home Management:** Controlling smart devices in your home, like lights and thermostats, or setting up automated routines with your voice.
- ☐ **Information and Calculations:** Asking for quick facts, performing calculations, or obtaining definitions and translations through spoken requests.
- ☐ **Navigation and Traffic:** Setting destinations and getting traffic updates for your commute or travel plans based on live traffic data.
- ☐ **Task Management:** Organizing your schedule, setting up appointments, and managing your to-do lists with voice commands.
- ☐ **Online Shopping Assistance:** Adding items to your shopping cart or list, placing orders for essentials, and tracking your purchases with your voice.
- ☐ **Voice-Enabled Financial Management:** Checking your bank account balance and making payments with voice commands
- ☐ **Voice controlling the vehicle:** Climate control, such as turning on and off fan or heating
- ☐ Other

6

What is it that you like about Voice Assistants?

7

What is it that you dislike about Voice Assistants?

8

How familiar are you with using the latest Conversational Chatbots (such as ChatGPT, Bing, Microsoft Copilot or Gemini)? *

- Never heard of it
- Heard about it, never used it
- Use it a few times a year
- Use it a few times a month
- Use it a few times a week
- Use it daily

9

When you use Conversational Chatbots, what do you use it for? *

- ☐ **Searching for information:** Asking a variety of questions, from general knowledge to specific queries about science, history, culture, and more.
- ☐ **Writing and editing:** Getting assistance with drafting, editing, and polishing written content, including emails, essays, reports, and creative writing, offering grammar corrections and stylistic suggestions.
- ☐ **Entertainment:** Getting help creating games, jokes, trivia, and interactive stories, or generating creative content like poems, stories, and song lyrics.
- ☐ **Social interaction:** Engaging in social interaction with the chatbot, having a conversation just for fun or chatting about everyday life.
- ☐ **Education and learning:** Receiving assistance with learning new topics, explaining complex subjects, and offering tutorials or step-by-step guides in areas like coding, mathematics, language learning, etc.
- ☐ **Productivity tasks:** Managing tasks and productivity, such as setting reminders, managing calendars, summarizing meetings or notes, and organizing to-do lists.
- ☐ **Programming help:** Obtaining assistance in coding, explaining programming concepts, debugging code, and receiving examples of code snippets in various programming languages.
- ☐ **News and Updates:** Receiving summaries of current news and updates on specific topics.
- ☐ **Travel and Navigation:** Looking for travel information, suggestions for itineraries, and advice on local attractions, accommodations, dining options and help navigating.
- ☐ Other

10

How reliable do you think Conversational Chatbots' answers are?
(Works consistently, understands and answers your questions accurately)

If you don't know, skip this question

1	2	3	4	5

Unreliable — Reliable

11

How trustworthy do you think Conversational Chatbots' answers are?
(Handles your data in a safe way and provides ethical answers)

If you don't know, skip this question

1	2	3	4	5

Untrustworthy — Trustworthy

12

How useful do you think Conversational Chatbots are?

If you don't know, skip this question

1	2	3	4	5

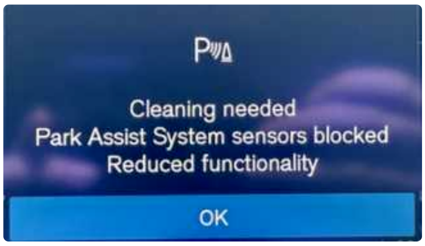
Useless — Useful

About your expectation

Imagine having an in-car voice assistant “Car Co-pilot” that knows about your car, can give support and help you report problems. In the following section, **think of the questions as if you had this “Car Co-pilot” voice assistant in your vehicle.**

You will be presented with a variety of scenarios related to your experiences while driving. Each scenario is designed to explore different aspects of using a Car Co-pilot system.

When I see a warning pop-up on the screen...



(such as the parking sensors need to be cleaned or that the airbag on the passenger side has been deactivated)

13

..I would like an **explanation** of what the warning pop-up indicates *

How important is it to get an explanation in this situation?

- ◯ Not at all important
- ◯ Somewhat important
- ◯ Moderately important
- ◯ Very important
- ◯ Extremely important

14

..I would like an **explanation** of what the warning pop-up indicates *

How likely is it that you would use the Car Co-pilot in this situation?

- ◯ Not at all likely
- ◯ Somewhat likely
- ◯ Moderately likely
- ◯ Very likely
- ◯ Extremely likely

15

..I would like to **report** this problem to Volvo *

How important is it for you to report the problem in this situation?

- Not at all important
- Somewhat important
- Moderately important
- Very important
- Extremely important

16

..I would like to **report** this problem to Volvo *

How likely is it that you would use the Car Co-pilot in this situation?

- Not at all likely
- Somewhat likely
- Moderately likely
- Very likely
- Extremely likely

17

..I would like to know how to **solve** it (such as guidance or troubleshooting) *

How important is it to get support with solving the problem in this situation?

- Not at all important
- Somewhat important
- Moderately important
- Very important
- Extremely important

18

..I would like to know how to **solve** it (such as guidance or troubleshooting) *
How likely is it that you would use the Car Co-pilot in this situation?

- Not at all likely
- Somewhat likely
- Moderately likely
- Very likely
- Extremely likely

When I see a warning light showing on the dash-board of my car...

19

..I would like an **explanation** of what the light indicates *

How important is it to get an explanation in this situation?

- ◯ Not at all important
- ◯ Somewhat important
- ◯ Moderately important
- ◯ Very important
- ◯ Extremely important

20

..I would like an **explanation** of what the light indicates *

How likely is it that you would use the Car Co-pilot in this situation?

- ◯ Not at all likely
- ◯ Somewhat likely
- ◯ Moderately likely
- ◯ Very likely
- ◯ Extremely likely

21

..I would like to **report** this problem to Volvo *

How important is it for you to report the problem in this situation?

- ◯ Not at all important
- ◯ Somewhat important
- ◯ Moderately important
- ◯ Very important
- ◯ Extremely important

22

..I would like to **report** this problem to Volvo *

How likely is it that you would use the Car Co-pilot in this situation?

- ◯ Not at all likely
- ◯ Somewhat likely
- ◯ Moderately likely
- ◯ Very likely
- ◯ Extremely likely

23

..I would like to know how to **solve** it (such as guidance or troubleshooting) *

How important is it to get support with solving the problem in this situation?

- ◯ Not at all important
- ◯ Somewhat important
- ◯ Moderately important
- ◯ Very important
- ◯ Extremely important

24

..I would like to know how to **solve** it (such as guidance or troubleshooting) *

How likely is it that you would use the Car Co-pilot in this situation?

- ◯ Not at all likely
- ◯ Somewhat likely
- ◯ Moderately likely
- ◯ Very likely
- ◯ Extremely likely

When I get a notification from the vehicle...

(such as the tires are worn out or are at incorrect pressures)

25

..I would like an **explanation** of what the problem is *
How important is it to get an explanation in this situation?

- Not at all important
- Somewhat important
- Moderately important
- Very important
- Extremely important

26

..I would like an **explanation** of what the problem is *
How likely is it that you would use the Car Co-pilot in this situation?

- Not at all likely
- Somewhat likely
- Moderately likely
- Very likely
- Extremely likely

27

..I would like to **report** this problem to Volvo *

How important is it for you to report the problem in this situation?

- Not at all important
- Somewhat important
- Moderately important
- Very important
- Extremely important

28

..I would like to **report** this problem to Volvo *

How likely is it that you would use the Car Co-pilot in this situation?

- Not at all likely
- Somewhat likely
- Moderately likely
- Very likely
- Extremely likely

29

..I would like to know how to **solve** it (such as guidance) *

How important is it to get support with solving the problem in this situation?

- Not at all important
- Somewhat important
- Moderately important
- Very important
- Extremely important

30

..I would like to know how to **solve** it (such as guidance) *

How likely is it that you would use the Car Co-pilot in this situation?

- Not at all likely
- Somewhat likely
- Moderately likely
- Very likely
- Extremely likely

When I notice something unusual with my car...

(such as odd noises, vibrations, or changes in my car's handling)

31

..I would like an **explanation** of what the problem is *

How important is it to get an explanation in this situation?

- ◯ Not at all important
- ◯ Somewhat important
- ◯ Moderately important
- ◯ Very important
- ◯ Extremely important

32

..I would like an **explanation** of what the problem is *

How likely is it that you would use the Car Co-pilot in this situation?

- ◯ Not at all likely
- ◯ Somewhat likely
- ◯ Moderately likely
- ◯ Very likely
- ◯ Extremely likely

33

..I would like to **report** this problem to Volvo *

How important is it for you to report the problem in this situation?

- Not at all important
- Somewhat important
- Moderately important
- Very important
- Extremely important

34

..I would like to **report** this problem to Volvo *

How likely is it that you would use the Car Co-pilot in this situation?

- Not at all likely
- Somewhat likely
- Moderately likely
- Very likely
- Extremely likely

35

..I would like to know how to **solve** it (such as guidance or troubleshooting) *

How important is it to get support with solving the problem in this situation?

- Not at all important
- Somewhat important
- Moderately important
- Very important
- Extremely important

36

..I would like to know how to **solve** it (such as guidance or troubleshooting) *

How likely is it that you would use the Car Co-pilot in this situation?

- Not at all likely
- Somewhat likely
- Moderately likely
- Very likely
- Extremely likely

When a function isn't working as expected...

(such as repeatedly losing Bluetooth connectivity or the steering wheel heating defaults to an unwanted level)

37

..I would like an **explanation** of what the problem is *

How important is it to get an explanation in this situation?

- ◯ Not at all important
- ◯ Somewhat important
- ◯ Moderately important
- ◯ Very important
- ◯ Extremely important

38

..I would like an **explanation** of what the problem is *

How likely is it that you would use the Car Co-pilot in this situation?

- ◯ Not at all likely
- ◯ Somewhat likely
- ◯ Moderately likely
- ◯ Very likely
- ◯ Extremely likely

39

..I would like to **report** this problem to Volvo *

How important is it for you to report the problem in this situation?

- ◯ Not at all important
- ◯ Somewhat important
- ◯ Moderately important
- ◯ Very important
- ◯ Extremely important

40

..I would like to **report** this problem to Volvo *
How likely is it that you would use the Car Co-pilot in this situation?

- Not at all likely
- Somewhat likely
- Moderately likely
- Very likely
- Extremely likely

41

..I would like to know how to **solve** it (such as guidance or troubleshooting) *
How important is it to get support with solving the problem in this situation?

- Not at all important
- Somewhat important
- Moderately important
- Very important
- Extremely important

42

..I would like to know how to **solve** it (such as guidance or troubleshooting) *
How likely is it that you would use the Car Co-pilot in this situation?

- Not at all likely
- Somewhat likely
- Moderately likely
- Very likely
- Extremely likely

When a new over the air (OTA) software update has been installed in my car...

43

..I would like to **give feedback** to Volvo about it *

How important is it for you to give feedback in this situation?

- Not at all important
- Somewhat important
- Moderately important
- Very important
- Extremely important

44

..I would like to **give feedback** to Volvo about it *

How likely is it that you would use the Car Co-pilot in this situation?

- Not at all likely
- Somewhat likely
- Moderately likely
- Very likely
- Extremely likely

When I have an idea about a car function or feature...

45

..I would like to **share** it with Volvo. *

How important is it for you to share your idea in this situation?

- ◯ Not at all important
- ◯ Somewhat important
- ◯ Moderately important
- ◯ Very important
- ◯ Extremely important

46

...I would like to **share** it with Volvo. *

How likely is it that you would use the Car Co-pilot in this situation?

- ◯ Not at all likely
- ◯ Somewhat likely
- ◯ Moderately likely
- ◯ Very likely
- ◯ Extremely likely

When I have a great car experience...

47

..I would like to **inform** Volvo *
How important is it for you to inform Volvo in this situation?

- Not at all important
- Somewhat important
- Moderately important
- Very important
- Extremely important

48

..I would like to **inform** Volvo *
How likely is it that you would use the Car Co-pilot in this situation?

- Not at all likely
- Somewhat likely
- Moderately likely
- Very likely
- Extremely likely

When I have trouble finding how to perform a task...

(such as setting up a new car profile or connecting my phone through Bluetooth)

49

..I would like to **report** this problem to Volvo *

How important is it for you to report the problem in this situation?

- Not at all important
- Somewhat important
- Moderately important
- Very important
- Extremely important

50

..I would like to **report** this problem to Volvo *

How likely is it that you would use the Car Co-pilot in this situation?

- Not at all likely
- Somewhat likely
- Moderately likely
- Very likely
- Extremely likely

51

..I would like to know how to **solve** it (such as guidance) *

How important is it to get support with solving the problem in this situation?

- Not at all important
- Somewhat important
- Moderately important
- Very important
- Extremely important

52

..I would like to know how to **solve** it (such as guidance) *

How likely is it that you would use the Car Co-pilot in this situation?

- Not at all likely
- Somewhat likely
- Moderately likely
- Very likely
- Extremely likely

When I struggle with understanding a function or how to use it...

(such as how to calibrate appropriate tire pressure, how to use Pilot Assist or what affects the range of an electric vehicle)

53

..I would like an **explanation** of the function *

How important is it to get an explanation in this situation?

- Not at all important
- Somewhat important
- Moderately important
- Very important
- Extremely important

54

..I would like an **explanation** of the function *

How likely is it that you would use the Car Co-pilot in this situation?

- Not at all likely
- Somewhat likely
- Moderately likely
- Very likely
- Extremely likely

55

..I would like to **report** this issue to Volvo *

How important is it for you to report this issue in this situation?

- Not at all important
- Somewhat important
- Moderately important
- Very important
- Extremely important

56

..I would like to **report** this issue to Volvo *

How likely is it that you would use the Car Co-pilot in this situation?

- ◯ Not at all likely
- ◯ Somewhat likely
- ◯ Moderately likely
- ◯ Very likely
- ◯ Extremely likely

Car Co-pilot initiating

Car Co-pilot can also initiate a conversation with you. For example, give proactive feedback on aspects related to your journey. For each of the statements below, rate the appropriateness for using the Car Co-pilot in this way.

57

The Car Co-pilot asks me about my opinion on the new Over The Air (OTA) software update that has been installed in my car. *

1	2	3	4	5

Not Appropriate Very Appropriate

58

The Car Co-pilot informs me about weather conditions (such as cold snaps or heat waves) that could affect my car or my driving, along with tips for protecting it *

1	2	3	4	5

Not Appropriate Very Appropriate

59

The Car Co-pilot provides me with information and tips about undiscovered car functionalities that may be suitable for me *

1	2	3	4	5

Not Appropriate Very Appropriate

Other

60

Do you think that you would give more feedback, ask questions or report problems if you could do so using a voice assistant, like a Car Co-pilot, while driving? *

- ◯ Not at all likely
- ◯ Somewhat likely
- ◯ Moderately likely
- ◯ Very likely
- ◯ Extremely likely

61

Are there any other situations where you would find the Car Co-pilot useful?

62

Do you have any concerns using the Car Co-pilot?

63

Anything else you would like to share regarding this questionnaire?

Following-up interview invitation

Thank you for completing our survey! Your input is valuable to us.

We are conducting a follow-up study to dig deeper into some of the topics covered in this survey. This involves a semi-structured interview lasting approximately 1 hour, which may include practical tests and observation sessions to better understand your experiences and perspectives.

64

Would you be interested in participating in a follow-up interview? *

◯ Yes, I am interested

◯ No, thank you

B

Survey Code Book

This appendix present a code book of the survey questions. This section contains a list of the full text for each question along with the abbreviations used throughout the report. Each scenario question i labeled with 'Q' followed by a number, as also presented in the Scatter Plots in the result chapter.

Background questions

Gender = *What is your gender?*
Age = *How old are you?*
Country = *Where do you currently live?*
VA_Fam = *How familiar are you with using voice Assistants*
VA_ Use = *What do you use Voice assistants for?*
VA_Like = *What is it that you like about Voice Assistants?*
VA_Dislike = *What is it that you dislike about Voice Assistants?*
CB_Fam = *How familiar are you with using the latest Conversational Chatbots (such as ChatGPT, Bing, Microsoft Copilot or Gemini)?*
CB_Use = *When you use Conversational Chatbots, what do you use it for?*
CB_Reli = *How reliable do you think Conversational Chatbots' answers are? (Works consistently, understands and answers your questions accurately)*
CB_Trust = *How trustworthy do you think Conversational Chatbots' answers are?*
(Handles your data in a safe way and provides ethical answers)
CB_Useful = *How useful do you think Conversational Chatbots are?*

Scenarios questions

Q1

When I see a warning Pop-up on the screen, I would like an explanation of what the warning pop-up indicates
Pop-up_Explanation_Importance = *How important is it to get an explanation in this situation?*
Pop-up_Explanation_Likely = *How likely is it that you would use the Car Co-pilot in this situation?*

When I see a warning Pop-up on the screen, I would like to report this problem to Volvo
Pop-up_Report_Importance = *How important is it for you to report the problem in this situation?*
Pop-up_Report_Likely = *How likely is it that you would use the Car Co-pilot in this situation?*

When I see a warning Pop-up on the screen, I would like to know how to solve it (such as help guidance troubleshooting)
Pop-up_Solve_Importance= *How important is it to get support with solving the problem in this situation?*
Pop-up_Solve_Likely = *How likely is it that you would use the Car Co-pilot in this situation?*

Q2

When I see a warning light showing in the dashboard of my car, I would like an explanation of what the light indicates

Light_Explanation_Importance = *How important is it to get an explanation in this situation?*
Light_Explanation_Likely = *How likely is it that you would use the Car Co-pilot in this situation?*

When I see a warning light showing in the dashboard of my car, I would like to report this problem to Volvo
Light_Report_Importance = *How important is it for you to report the problem in this situation?*
Light indicates+Report+Likely = *How likely is it that you would use the Car Co-pilot in this situation?*

When I see a warning light showing in the dashboard of my car, I would like to know how to solve it (such as help guidance troubleshooting)
Light_Solve_Importance = *How important is it to get support with solving the problem in this situation?*
Light_Solve_Likely = *How likely is it that you would use the Car Co-pilot in this situation?*

Q3

When I get a notification from the vehicle, I would like an explanation of what the problem is
Notification_Explanation_Importance = *How important is it to get an explanation in this situation?*
Notification_Explanation_Likely = *How likely is it that you would use the Car Co-pilot in this situation?*

When I get a notification from the vehicle, I would like to report this problem to Volvo
Notification_Report_Importance = *How important is it for you to report the problem in this situation?*
Notification_Report_Likely = *How likely is it that you would use the Car Co-pilot in this situation?*

When I get a notification from the vehicle, I would like to know how to solve it (such as help guidance troubleshooting)
Notification_Solve_Importance = *How important is it to get support with solving the problem in this situation?*
Notification_Solve_Likely = *How likely is it that you would use the Car Co-pilot in this situation?*

Q4

When I notice something unusual with my car, I would like an explanation of what the problem is
Unusual_Explanation_Importance = *How important is it to get an explanation in this situation?*
Unusual_Explanation_Likely = *How likely is it that you would use the Car Co-pilot in this situation?*

When I notice something unusual with my car, I would like to report this problem to Volvo
Unusual_Report_Importance = *How important is it for you to report the problem in this situation?*
Unusual_Report_Likely = *How likely is it that you would use the Car Co-pilot in this situation?*

When I notice something unusual with my car, I would like to know how to solve it (such as help guidance troubleshooting)
Unusual_Solve_Importance = *How important is it to get support with solving the problem in this situation?*
Unusual_Solve_Likely = *How likely is it that you would use the Car Co-pilot in this situation?*

Q5

When a function isn't working as expected, I would like an explanation of what the problem is
Function_Explanation_Importance = *How important is it to get an explanation in this situation?*
Function_Explanation_Likely= *How likely is it that you would use the Car Co-pilot in this situation?*

When a function isn't working as expected, I would like to report this problem to Volvo
Function_Report_Importance = *How important is it for you to report the problem in this situation?*
Function_Report_Likely = *How likely is it that you would use the Car Co-pilot in this situation?*

When a function isn't working as expected, I would like to know how to solve it (such as help guidance troubleshooting)
Function_Solve_Importance = *How important is it to get support with solving the problem in this situation?*
Function_Solve_Likely = *How likely is it that you would use the Car Co-pilot in this situation?*

Q6

When I have trouble finding how to perform a task, I would like to report this problem to Volvo
Find_Report_Importance = *How important is it for you to report the problem in this situation?*
Find_Report_Likely = *How likely is it that you would use the Car Co-pilot in this situation?*

When I have trouble finding how to perform a task, I would like to know how to solve it (such as help guidance troubleshooting)
Find_Solve_Importance = *How important is it to get support with solving the problem in this situation?*
Find_Solve_Likely = *How likely is it that you would use the Car Co-pilot in this situation?*

Q7

<u>When I struggle with understanding a function or how to use it, I would like an explanation of the function</u>
UnderstandFunc_Explanation_Importance = *How important is it to get an explanation in this situation?*
UnderstandFunc_Explanation_Likely = *How likely is it that you would use the Car Co-pilot in this situation?*

<u>When I struggle with understanding a function or how to use it, I would like to report this issue to Volvo</u>
Understanding a function or how to use it+Report+Importance = *How important is it for you to report the issue in this situation?*
Understanding a function or how to use it+Report+Likely = *How likely is it that you would use the Car Co-pilot in this situation?*

Feedback

<u>When a new over the air (OTA) software update has been installed in my car, I would like to give feedback to Volvo about it</u>
OTA_Feedback_Importance = *How important is it for you to give feedback in this situation?*
OTA_Feedback_Likely = *How likely is it that you would use the Car Co-pilot in this situation?*

Share

<u>When I have an idea about a function or feature, I would like to share it with Volvo.</u>
Idea_Share_Importance = *How important is it for you to share your idea with Volvo in this situation?*
Idea_Share_Likely = *How likely is it that you would use the Car Co-pilot in this situation?*

Inform

<u>When I have a great car experience, I would like to inform Volvo</u>
Experience_Inform_Importance = *How important is it for you to inform Volvo in this situation?*
Experience_Inform_Likely = *How likely is it that you would use the Car Co-pilot in this situation?*

Car Co-pilot initiating interaction questions

VA_Initiate_OTA = *The Car Co-pilot asks me about my opinion on the new Over The Air (OTA) software update that has been installed in my car.*

VA_Initiate_Weather = *The Car Co-pilot informs me about weather conditions (such as cold snaps or heat waves) that could affect my car or my driving, along with tips for protecting it*

VA_Initiate_Info = *The Car Co-pilot provides me with information and tips about undiscovered car functionalities that may be suitable for me*

Finishing questions

VA_More_Feedback = *Do you think that you would give more feedback, ask questions or report problems if you could do so using a voice assistant, like a Car Co-pilot, while driving?*

VA_Usefulness = *Are there any other situations where you would find the Car Co-pilot useful?*

VA Concerns = *Do you have any concerns using the Car Co-pilot?*

VA_Other = *Anything else you would like to share regarding this questionnaire?*

C

Expert Interview Questions

This appendix presents the questions that was asked during the expert interviews.

Background questions

- How is what you do or have done related to Voice Assistants?
- What is your experience with voice assistants?
  - Do you have them in your home or in your car (if you have one)? Google assistant, Amazon Alexa, Siri for example
  - What do you use them for?
  - In your opinion, what do you think about them?
- Based on your driving habits and vehicle use, what specific functionalities do you seek from an in-car voice assistant?

Current state of voice assistants

- Can you describe the current state of voice and dialogue systems in vehicles?
- Can you describe the current state of voice assistants for personal use such as at home or on the phone?
- Have you noticed any limitations with current voice assistants in terms of understanding commands or providing relevant responses? How do you think generative AI, such as ChatGPT or Large language models could address these issues?
- What are the key features you are working on implementing right now? Why are they important? (Only for interviewee 2)

Competitor analysis (only for interviewee 3)

- What do you know about competitors and how they are using Voice Assistants in their vehicles?
- Is there any particular competitor that has integrated Voice Assistants in a successful way (What is the State of the Art)? If so, what made them successful?
- Can you provide insights into any standout features or strategies used by competitors?
- Have any of these competitors been/are working on integrating Generative AI into their systems?

Customer expectations

- Based on your experience, what are the key customer needs and expectations from voice assistants in vehicles?
- In what ways do you believe AI can further enhance the user experience with voice assistants in cars? Are there any specific AI capabilities that would meet customer expectations?
- What are the opportunities with using Generative AI in voice assistants in vehicles?
- What are the technological challenges with using Generative AI in voice assistants in vehicles?
- From a technological viewpoint, do you think these dialogue systems in vehicles could be improved with Generative AI? How and why?

Concerns

- Do you have any concerns using AI-enabled voice assistants in vehicles?