Universität zu Köln

Data-Driven Evaluation of In-Vehicle Information Systems

Inaugural-Dissertation zur Erlangung des Doktorgrades der Mathematisch-Naturwissenschaftlichen Fakultät der Universität zu Köln

vorgelegt von Patrick Ebel aus Menden

Köln, 2023

Berichterstattende (Gutachtende):

1. Prof. Dr. Andreas Vogelsang, Universität zu Köln
2. Prof. Dr. Martin Baumann, Universität Ulm
3. Prof. Dr. Nele Rußwinkel, Universität Lübeck

Tag der mündlichen Prüfung: 10.07.2023

Permission to use the cover image granted by Mercedes-Benz Group Media. Source: Fenzl et al. [9]

Abstract

Today's In-Vehicle Information Systems (IVISs) are feature-rich systems that provide the driver with numerous options for entertainment, information, comfort, and communication. Drivers can stream their favorite songs, read reviews of nearby restaurants, or change the ambient lighting to their liking. To do so, they interact with large center stack touchscreens that have become the main interface between the driver and IVISs. To interact with these systems, drivers must take their eyes off the road which can impair their driving performance. This makes IVIS evaluation critical not only to meet customer needs but also to ensure road safety. The growing number of features, the distraction caused by large touchscreens, and the impact of driving automation on driver behavior pose significant challenges for the design and evaluation of IVISs. Traditionally, IVISs are evaluated qualitatively or through small-scale user studies using driving simulators. However, these methods are not scalable to the growing number of features and the variety of driving scenarios that influence driver interaction behavior. We argue that data-driven methods can be a viable solution to these challenges and can assist automotive User Experience (UX) experts in evaluating IVISs. Therefore, we need to understand how data-driven methods can facilitate the design and evaluation of IVISs, how large amounts of usage data need to be visualized, and how drivers allocate their visual attention when interacting with center stack touchscreens.

In **Part I**, we present the results of two empirical studies and create a comprehensive understanding of the role that data-driven methods currently play in the automotive UX design process. We found that automotive UX experts face two main conflicts: First, results from qualitative or small-scale empirical studies are often not valued in the decision-making process. Second, UX experts often do not have access to customer data and lack the means and tools to analyze it appropriately. As a result, design decisions are often not user-centered and are based on subjective judgments rather than evidence-based customer insights. Our results show that automotive UX experts need data-driven methods that leverage large amounts of telematics data collected from customer vehicles. They need tools to help them visualize and analyze customer usage data and computational methods to automatically evaluate IVIS designs.

In **Part II**, we present ICEBOAT, an interactive user behavior analysis tool for automotive user interfaces. ICEBOAT processes interaction data, driving data, and glance data, collected over-the-air from customer vehicles and visualizes it on different levels of granularity. Leveraging our multi-level user behavior analysis framework, it enables UX experts to effectively and efficiently evaluate driver interactions with touchscreen-based IVISs concerning performance and safety-related metrics.

In **Part III**, we investigate drivers' multitasking behavior and visual attention allocation when interacting with center stack touchscreens while driving. We present the first naturalistic driving study to assess drivers' tactical and operational self-regulation with center stack touchscreens. Our results show significant differences in drivers' interaction and glance behavior in response to different levels of driving automation, vehicle speed, and road curvature. During automated driving, drivers perform more interactions per touchscreen sequence and increase the time spent looking at the center stack touchscreen. These results emphasize the importance of context-dependent driver distraction assessment of driver interactions with IVISs. Motivated by this we present a machine learning-based approach to predict and explain the visual demand of in-vehicle touchscreen inter-

actions based on customer data. By predicting the visual demand of yet unseen touchscreen interactions, our method lays the foundation for automated data-driven evaluation of early-stage IVIS prototypes. The local and global explanations provide additional insights into how design artifacts and driving context affect drivers' glance behavior.

Overall, this thesis identifies current shortcomings in the evaluation of IVISs and proposes novel solutions based on visual analytics and statistical and computational modeling that generate insights into driver interaction behavior and assist UX experts in making user-centered design decisions.

Acknowledgements

Many people have supported me over the years and deserve my sincere thanks. Without them, this thesis would not have been possible.

First and foremost, I would like to thank my supervisor **Andreas Vogelsang**, I could not have asked for a better person to guide this research endeavor. From my start as a working student to my master's thesis to my PhD, thank you for your unconditional support, thesis related or not. Thank you for challenging me, for believing in me, and for all that you have taught me along the way. You have taught me not only to do research but also how to enjoy it without sacrificing the life around me. You are a role model to me.

Thank you **Martin Baumann** for your valuable feedback and for serving as a second supervisor on this thesis. I hope this is just the beginning and that more collaborations will follow. The same goes for my third supervisor **Nele Rußwinkel**. Thank you!

Christoph Lingenfelder, my unofficial supervisor and one of the most knowledgeable individuals I've ever met. Thank you for all your ideas and constant support over the past years. Thank you for making this project possible.

Thanks to all my colleagues at **MBition**, **Mercedes-Benz**, and the **Chair of Software and Systems Engineering**, whether in Berlin, Cologne, Stuttgart, or Sunnyvale. There are too many to mention, but the research for this thesis would not have been possible without your support and inspiration. I'm grateful for the many hours we spent in (mostly virtual) meetings navigating the sea of obstacles we encountered along the way.

Some special gratitude goes to the **AutoUI Community** and the many individuals who have supported me throughout my journey. This community has been incredibly welcoming and it has been so easy for someone from the outside to get involved. From a rough start with two virtual conferences to sharing late-night conversations over beers on a rooftop in Seoul. This community is one of the main reasons I want to stay in academia. I couldn't say this without the person who introduced me to this great group of researchers, and who I'm lucky to call a friend: **Debargha (Dave) Dey**. You are one of the kindest, most considerate, and caring people I have met along the way. I owe you one, and I hope you still accept ice cream as payment. Speaking of people who have opened doors for me: Thank you **Chris Janssen** for your support and for inviting me to the Dagstuhl Seminar, which turned out to be one of the most inspiring weeks of my research journey so far. I would especially like to thank **Julia Orlovska**, **Melanie Berger**, and **Kim Julian Gülle**. Collaborating with you was a joy!

To **all my friends** in Berlin and everywhere else: Thank you for all the joy you bring into my life and for helping me maintain a healthy balance between work time and aperitivo time. I especially want to thank **Flo**. Thank you for always listening to me (and my rants), for the late-night reviews, and for always being there. It's a shame that our resumes are no longer exactly the same.

To my **parents**: Thank you for making this possible. Even though I was the first in our family to go to university, I never felt like I was at a disadvantage because of your support. Thank you **Tobi** for always checking in on me and being the caring little brother you are.

Finally, and most importantly, I want to thank my girlfriend, **Eli**, for her patience, understanding, and emotional support through the ups and especially the downs of my PhD. You've always been there for me, made me smile, and without you by my side, this wouldn't have been possible. Also, thank you for letting me have the good chair. I love you.

Publication Preface

The contribution of this thesis is largely based on the following previously published work:

[1] P. Ebel, F. Brokhausen, and A. Vogelsang, "The Role and Potentials of Field User Interaction Data in the Automotive UX Development Lifecycle: An Industry Perspective," in *Proceedings of the 12th International Conference on Automotive User Interfaces and Interactive Vehicular Applications.* Virtual Event DC USA: ACM, Sep. 2020, pp. 141–150

[2] P. Ebel, J. Orlovska, S. Hünemeyer, C. Wickman, A. Vogelsang, and R. Söderberg, "Automotive UX design and data-driven development: Narrowing the gap to support practitioners," *Transportation Research Interdisciplinary Perspectives*, vol. 11, p. 100455, Sep. 2021

[3] P. Ebel, C. Lingenfelder, and A. Vogelsang, "Visualizing Event Sequence Data for User Behavior Evaluation of In-Vehicle Information Systems," in *Proceedings of the 13th International Conference on Automotive User Interfaces and Interactive Vehicular Applications.* Leeds United Kingdom: ACM, Sep. 2021, pp. 219–229

[4] P. Ebel, M. Berger, C. Lingenfelder, and A. Vogelsang, "How Do Drivers Self-Regulate their Secondary Task Engagements? The Effect of Driving Automation on Touchscreen Interactions and Glance Behavior," in *Proceedings of the 14th International Conference on Automotive User Interfaces and Interactive Vehicular Applications.* Seoul Republic of Korea: ACM, Sep. 2022, pp. 263–273, **Honorable Mention Award**

[5] P. Ebel, K. J. Gülle, C. Lingenfelder, and A. Vogelsang, "ICEBOAT: An Interactive User Behavior Analysis Tool for Automotive User Interfaces," in *The Adjunct Publication of the 35th Annual ACM Symposium on User Interface Software and Technology*, Aug. 2022

[6] P. Ebel, C. Lingenfelder, and A. Vogelsang, "On the forces of driver distraction: Explainable predictions for the visual demand of in-vehicle touchscreen interactions," *Accident Analysis & Prevention*, vol. 183, p. 106956, Apr. 2023, **HFES Europe Early Career Best Paper Award 2023**

[7] P. Ebel, C. Lingenfelder, and A. Vogelsang, "Multitasking while driving: How drivers self-regulate their interaction with in-vehicle touchscreens in automated driving," *International Journal of Human–Computer Interaction*, pp. 1–18, 2023, *Extended version of [4]*

[8] P. Ebel, K. J. Gülle, C. Lingenfelder, and A. Vogelsang, "Exploring Millions of User Interactions with ICEBOAT: Big Data Analytics for Automotive User Interfaces," in *Proceedings of the 15th International Conference on Automotive User Interfaces and Interactive Vehicular Applications*, Ingolstadt, Germany, 2023*, accepted for publication*

Contribution Statement

Patrick Ebel is the lead author of all publications included in this thesis. As lead author, he took primary responsibility for the design, implementation, data collection and analysis, and publication of the results in peer-reviewed venues. His and the co-authors' contributions to the included publications are described below using the Contributor Roles Taxonomy (CRediT):

[1]: **Patrick Ebel:** Conceptualization, Methodology, Data Curation, Formal Analysis, Investigation, Writing – Original Draft. **Florian Brokhausen:** Conceptualization, Methodology, Data Curation, Formal Analysis, Investigation, Writing – Original Draft. **Andreas Vogelsang:** Supervision, Writing – Review & Editing.

[2]: **Patrick Ebel:** Conceptualization, Methodology, Data Curation, Formal Analysis, Writing – Original Draft. **Julia Orlovska:** Conceptualization, Methodology, Data Curation, Formal Analysis, Writing – Original Draft. **Sebastian Hünemeyer:** Conceptualization, Writing – Original Draft. **Casper Wickman:** Supervision, Writing – Review & Editing. **Andreas Vogelsang:** Supervision, Writing – Review & Editing. **Rikard Söderberg:** Supervision, Writing – Review & Editing.

[3]: **Patrick Ebel:** Conceptualization, Methodology, Data Curation, Software, Visualization, Writing – Original Draft. **Christoph Lingenfelder:** Resources, Supervision, Writing – Review & Editing. **Andreas Vogelsang:** Supervision, Writing – Review & Editing.

[4]: **Patrick Ebel:** Conceptualization, Methodology, Data Curation, Formal Analysis, Visualization, Writing – Original Draft. **Moritz Berger:** Writing – Review & Editing. **Christoph Lingenfelder:** Resources, Supervision. **Andreas Vogelsang:** Supervision, Writing – Review & Editing.

[5]: **Patrick Ebel:** Conceptualization, Methodology, Investigation, Data Curation, Formal Analysis, Visualization, Writing – Original Draft. **Kim Julian Gülle:** Investigation, Software, Visualization. **Christoph Lingenfelder:** Resources, Supervision. **Andreas Vogelsang:** Supervision, Writing – Review & Editing.

[6]: **Patrick Ebel:** Conceptualization, Methodology, Data Curation, Formal Analysis, Visualization, Writing – Original Draft. **Christoph Lingenfelder:** Resources, Supervision, Writing – Review & Editing. **Andreas Vogelsang:** Supervision, Writing – Review & Editing.

[7]: **Patrick Ebel:** Conceptualization, Methodology, Data Curation, Formal Analysis, Visualization, Writing – Original Draft. **Christoph Lingenfelder:** Resources, Supervision. **Andreas Vogelsang:** Supervision, Writing – Review & Editing.

[8]: **Patrick Ebel:** Conceptualization, Methodology, Investigation, Data Curation, Formal Analysis, Visualization, Writing – Original Draft. **Kim Julian Gülle:** Investigation, Software, Visualization. **Christoph Lingenfelder:** Resources, Supervision. **Andreas Vogelsang:** Supervision, Writing – Review & Editing.

Contents

List of Figures

List of Tables

List of Acronyms

ACC Advanced Cruise Control

ACT-R Adaptive Control of Thought—Rational

ADAS Advanced Driver Assistance System

AOI Area of Interest

API Application Programming Interface

CAN Controller Area Network

CDDB Component Description Data Base

CE Continuous Experimentation

CI Continuous Integration

CRediT Contributor Roles Taxonomy

ECU Electronic Control Unit

ETL Extract, Transform, Load

FNN Feedforward Neural Network

GDPR General Data Protection Regulation

HCI Human-Computer Interaction

HMI Human-Machine Interface

HU Head Unit

HUD Head-Up Display

IVIS In-Vehicle Information System

JSON JavaScript Object Notation

KLM Keystroke-Level Model

KPI Key Performance Indicator

LCA Lane Centering Assist

LIME Local Interpretable Model-Agnostic Explanations

MAE Mean Absolute Error

NHTSA National Highway Safety Traffic Association

OEM Original Equipment Manufacturer

OTA Over-The-Air

PD Product Development

POI Point of Interest

RH Research Hub

SD Standard Deviation

SHAP SHapley Additive exPlanation

SHRP2 Strategic Highway Research Program 2

SOP Start of Production

SUS System Usability Scale

UCD User-Centered Design

UI User Interface

UX User Experience

XAI Explainable AI

Chapter 1

Introduction

This thesis addresses the question how data-driven methods can facilitate the design and evaluation of IVISs. First, we present current trends and challenges in the automotive domain and their impact on drivers' interactions with IVISs (see Section 1.1). In Section 1.2 we explain why these trends pose significant challenges for the user-centered design and evaluation of IVISs and how this motivates the search for data-driven methods that complement and improve the automotive UX design process. In Section 1.3 we present the main contributions of this thesis. Finally, Section 1.4 gives an overview of the structure of the thesis.

1.1 In-Vehicle Information Systems: Trends and Challenges

IVISs are menu-based systems that provide information and entertainment to the driver. Their main purpose is to enhance the driving experience [10]. They can include a variety of functions such as navigation, communication (e.g., phone or text messaging), entertainment (e.g., streaming music or video), or comfort (e.g., climate control or ambient lighting). The controls of IVISs are often integrated into a single screen-based device [11], and drivers can interact with them using different modalities such as hardkeys, touchscreens, gestures, or speech.

Nowadays, the design of IVISs has crucial impact on the overall user experience of a vehicle and customers expect high levels of usability and functionality from these systems [12]. However, the design and evaluation of IVISs is influenced, guided, and constrained by several factors. These include ever-increasing customer demands but also safety guidelines and regulations [13, 14]. To offer the best possible experience, it is therefore important to thoroughly evaluate how drivers interact with the IVISs. Thus, Original Equipment Manufacturers (OEMs) do not only need to balance the constantly growing demand for technology with the users' needs for a usable Human-Machine Interface (HMI) [11], but also need to consider safety implications that interactions with IVISs might have to driving [15, 16]. This balancing act is strongly influenced by the following three challenges:

Challenge 1: The Number of Features is Growing Digital products are evolving rapidly. Their impact on our daily lives is growing quickly as they become smarter, more capable, and incorporate more and more features. From tablets to smartwatches to smart refrigerators, they all demand our attention and interaction. The trend toward systems that incorporate an ever-increasing amount of features equally applies to the automotive industry and in particular to IVISs [12]. Originally, cars were purely manual products and transportation their only purpose. Later, simple IVISs in form of radios and rudimentary center stack consoles were introduced [11]. These systems allowed drivers to perform secondary

Figure 1.1: The Mercedes-Benz multi media system consisting of a large center stack touchscreen and an optional passenger display [9].

tasks, such as setting the radio station or route guidance, in addition to the primary driving task. However, customer demands for more functionality led to a fast increase in complexity [11] such that today's IVISs in addition to controlling vehicle, navigation, and comfort functions, incorporate various infotainment options from music and television to in-car video games. As a results, modern IVISs are highly advanced and fully connected systems (see Figure 1.1). Interacting with them is not much different from interacting with modern smartphones or tablets and the increasing number of features makes it difficult to evaluate modern IVISs with traditional qualitative and resource-intensive methods.

Challenge 2: Touchscreens are Becoming the Main Interface Between Driver and IVIS

Although the first touchscreen-based IVISs made their way into cars in the late 1990s [11], most functions were still controlled by haptic buttons or alternative controls such as rotary knobs or centrally located touchpads. However, after the introduction of the iPhone in 2007 [17] and its subsequent market penetration, touchscreens of all sizes quickly became the de facto interface for smart devices. They became ubiquitous, not only in our personal devices, but also in public spaces. Today, large center stack touchscreens, such as those found in Tesla's Model 3[1] or the Mercedes-Benz EQS[2], are the primary interface between the driver and the IVISs. With this development, haptic buttons are gradually disappearing from today's vehicles. However, the use of touchscreens and the wide range of functions now available pose new challenges in terms of driver distraction and road safety. First, touchscreen interfaces require more visual attention than haptic interfaces. Without the tactile feedback provided by haptic interfaces, users must visually sample the interface

[1] https://www.tesla.com/model3

[2] https://www.mercedes-benz.com/en/innovation/future-mobility/eqs-with-unique-mbux-hyperscreen/

and verify that they made the correct selection [18]. This causes drivers to take their eyes off the road for longer periods of time. This may not be a problem for smartphones or computers, where interaction with the device is the primary task. However, during manual or partially-automated driving (Level 1 and Level 2 according SAE J3016 [19]), the driver must constantly monitor the driving environment. Consequently, interactions with IVISs are considered a secondary task and time spent looking away from the road is directly correlated with increased crash risk [16].

Challenge 3: Driving Automation Impacts Drivers' Interaction Behavior The primary goal of automated driving features such as Advanced Cruise Control (ACC) and Lane Centering Assist (LCA) is not only to make driving more comfortable, but also to make driving safer. Several studies show that these systems can make driving safer and can reduce the incidence of critical situations [20, 21]. However, even though automated driving features are more widely available and more capable than ever, the number of crashes due to human error caused by distraction has stagnated in recent years [14]. Driving automation not only has a positive impact on driving safety, but also tends to increase the margins in which drivers consider it safe to engage in non-driving-related tasks [22, 23, 24]. To interact with IVISs while driving, drivers must divide their attention between the primary driving task and the non-driving-related secondary task. Although drivers have been shown to self-regulate their engagement in secondary tasks based on driving demands [25, 26, 27], this task-switching behavior is critical. In addition, drivers tend to overestimate the capabilities of automated driving functions [28]. This can increase the likelihood that they will engage in non-driving related tasks in situations where they should be constantly monitoring the driving environment [29, 30]. For this reason, it is particularly important to consider the influence of driving automation on the interaction between the driver and the IVIS.

How does this Affect the Design and Evaluation of IVISs? The increasing number of features, the introduction of large touchscreens, and the continuous evolution of Advanced Driver Assistance Systems (ADASs) are significantly changing how drivers interact with IVISs and how they experience these interactions. Furthermore, IVISs are not only compared among OEMs, but the user experience is shaped and influenced by the experiences that customers have with smartphones, tablets and websites. These developments make it increasingly important to design IVISs around the needs of the user. Thus, the automotive industry gradually moves from a traditional technology-driven development approach to a more user-centered one [31]. User-Centered Design (UCD) is an iterative, multidisciplinary design approach in which designers involve users at every stage of the process. The integration of users and their needs throughout the design process is considered essential to create a product with good usefulness and usability [32]. However, the continuous involvement of users and the need for experienced designers make UCD an expensive task. Combined with the aforementioned challenges, the effort associated with a traditional, purely qualitative UX approach often exceeds the resources of OEMs.

1.2 Problem Statement

Companies that work in digital domains such as web or app development are already enhancing their UX design and evaluation processes by integrating data-driven methods. These methods rely on large amounts of usage data collected at runtime to provide fast and objective user feedback. For example, modern websites can track every click of every user, resulting in large amounts of data that allow UX experts to quickly gain insights into user behavior and interests [33]. Knowing where people click, how long they interact with the system, and what they end up buying helps companies to tailor their services to customer needs. The triangulation of traditional, mostly qualitative, research methods with data-driven approaches diversifies user research which can lead to unexpected and more reliable insights [34]. We found that these data-driven methods are not yet leveraged in the automotive domain [1, 2]. IVIS evaluation still relies mostly on feedback from small-scale user studies. These often require fully functional prototypes, detailed test protocols, and expensive instrumentation. While experiments in high-fidelity driving simulators or on test tracks can provide valid results, they are not scalable to the many dimensions that influence driver behavior and experience, nor to the rapid evolution of vehicle technology. Furthermore, throughout the UX design process, evidence-based design decisions derived from qualitative or small-scale user studies are often overruled by the subjective opinions of decision makers [1]. This practice contradicts the fundamental principle of user centered design to actively involve users in product design and development [32]. Consequently, it can lead to oversimplifications with potentially severe usability and safety implications. This becomes evident when considering that the usability of infotainment systems has been the biggest source of problems for new car owners for several years [35, 36, 37].

To maximize the benefits of IVISs while keeping the risks associated with their distraction potential low, the design and evaluation of these systems must be user-centered. UX experts formulate the demand for methods, that allow them to quickly and continuously assess how users interact with IVISs [1]. They want to know how design affects usability and driver distraction, and how drivers adapt their interaction behavior based on the driving context. We argue that the use of data-driven methods becomes increasingly beneficial to evaluate IVISs for usability and safety and to inform decision-making processes throughout the UX design process. Since data-driven methods and automated analyses based on usage data from customers are easily scalable they can be a useful complement to current, largely qualitative approaches. Consequently, we formulate the following **problem statement**:

> The growing number of features, the increasing use of touchscreens while driving, and the impact of driving automation on driver behavior pose significant challenges for the design and evaluation of IVISs. Traditionally, IVIS are evaluated qualitatively or based on data collected from user studies conducted in laboratory environments. These methods are neither scalable to the growing number of features offered by IVISs, nor to the variety of driving scenarios in which drivers interact with these systems. The effort associated with this traditional approach is far beyond what automotive OEMs can afford. As a result, many decisions in the automotive UX design process are not user-centered due to a lack of customer insight. This results in potentially distracting user interfaces that do not meet customer needs.

1.3 Contributions

To design IVISs that meet customer expectations and are safe to use, we need to understand how data-driven methods can facilitate the design and evaluation of IVIS, how large amounts of data need to be visualized, and how drivers allocate their visual attention when interacting with center stack touchscreens. Accordingly, this thesis is guided by three complementary research questions:

RQ1: How can data-driven methods facilitate the design and evaluation of IVISs?

RQ2: How to visualize large amounts of data to effectively and efficiently analyze drivers' IVIS usage?

RQ3: How do drivers allocate their visual attention when interacting with center stack touchscreens while driving?

To answer these questions, we make four main contributions as listed below:

Contribution 1: Empirical Studies on IVIS Design and Data-Driven Methods

Published at:
AutoUI [1]
10 Pages
Full Paper
TRIP [2]
16 Pages
Full Paper

To lay the foundation for this dissertation, we conducted two empirical studies. The first study [1] provides a comprehensive overview of the role of data-driven methods in the design process of IVISs. We conducted interviews with 14 UX experts, 8 from automotive and 6 from digital companies, and analyzed the results using emergent thematic coding. We compare the current state-of-the-art in the automotive domain with digital domains, present the potentials practitioners see in analyzing user interaction data, and explore the concerns they share about data-driven methods. The second study [2] extends the initial study by an interview study conducted by Orlovska et al. [38] and two practical investigations on the use of data-driven methods in the product development process of OEMs. Our study highlights the main limitations that currently prevent the use of data-driven methods for the evaluation of IVISs. Motivated by these shortcomings, we identify the needs of UX experts for data-driven methods to improve user-centered and evidence-based decision-making in the UX design process. Based on these needs, we present potential use cases to improve IVIS design and evaluation, and formulate recommendations on how to better integrate data-driven methods into the UX design process. Both studies [1, 2] together provide a comprehensive understanding of the role, potentials, and limitations of applying data-driven methods in the automotive UX design process. These insights enable us to develop data-driven methods that enhance current practices and improve the quality of design decisions. Our key findings include:

- Insights based on qualitative and small-scale user studies are often not valued and resulting design decisions are overruled by management.
- UX experts need statistical support based on customer usage data to support design hypotheses, feature elicitation, and prioritization.
- Customer usage data is often unavailable or inaccessible throughout the design process due to organizational, legal, or technical constraints.
- UX experts need tools to analyze and visualize large amounts of customer data and data-driven methods for automated design evaluation.

Contribution 2: An Interactive User Behavior Analysis Tool for Automotive User Interfaces

Published at:
AutoUI [3]
11 Pages
Full Paper
UIST [5]
5 Pages
Short Paper
Accepted at:
AutoUI [8]
12 Pages
Full Paper

Our exploratory studies [1, 2] reveal that UX experts often lack access to customer data and tools for data analysis. To address this issue, we developed *ICEBOAT*, an interactive tool that processes and visualizes large amounts of telematics data collected from production vehicles. We developed ICEBOAT in a two-step process: First, we designed three visualizations that allow UX experts to explore user interaction data, driving data, and glance data on different levels of granularity [3]. Our user study shows that the visualizations assist UX experts in finding usability problems and unexpected user behavior. In a second step, using a co-design approach, we build on the initial visualization to develop ICEBOAT [8]. ICEBOAT allows users to easily define the tasks they want to analyze, automates data processing and visualization and connects the three levels of visualization through an interactive drill-down concept. Our usability and context of use evaluation (N=12) shows that ICEBOAT enables UX experts to efficiently generate knowledge that facilitates data-driven design decisions within the automotive UX design process. Our main contributions include:

- Semi-structured interviews to understand how big data visualizations need to be designed and integrated such that they can improve decision-making in the automotive UX design process.
- A multi-level user behavior visualization framework that provides effective visualizations of driver behavior at three levels of granularity.
- An interactive visual analytics tool that allows UX experts to effectively and efficiently evaluate touchscreen-based IVIS.

Contribution 3: Empirical Study on Drivers' Self-Regulation of Secondary Touchscreen Tasks

Published at:
AutoUI [4]
11 Pages
Full Paper
IJHCI [7]
23 Pages
Full Paper

We apply multilevel models to a real-world driving dataset to investigate how driving automation affects drivers' interactions with center stack touchscreens. We analyzed more than 30,000 secondary task engagements extracted from over 10,000 individual trips. More than 100 test vehicles contributed to the data collection from mid-October 2021 to mid-October 2022. Our results show that drivers self-regulate their behavior in response to changes in driving demand. We present significant differences in drivers' interaction and glance behavior in response to different levels of driving automation, vehicle speed, and road curvature. Our key findings are:

- Drivers are more likely to engage in secondary touchscreen tasks during partially automated driving. During level 2 automated driving, the mean glance duration toward the center stack touchscreen increases by 36 % and the mean number of interactions per sequence increases by 17 % compared to manual driving.
- During partially automated driving, driers use complex User Interface (UI) elements (e.g., maps) and touch gestures (e.g., multitouch) significantly more often.
- The effect of driving automation on drivers' self-regulation is larger than that of vehicle speed and road curvature.

Contribution 4: An Approach to Predict the Visual Demand of In-Vehicle Touchscreen Interactions

Published at:
AAP [6]
17 Pages
Full Paper

Motivated by UX experts' need for methods to automatically evaluate IVIS designs, we present a machine learning method that predicts the visual demand of in-vehicle touchscreen interactions based on expected usage scenarios. We use the SHapley Additive exPlanation (SHAP) method to provide local and global explanations of the factors influencing the model predictions. Enhanced with explanations, the predictions can help UX experts not only evaluate current IVIS designs, but also better anticipate and understand the impact of their design decisions on future designs. We evaluated the approach using a naturalistic driving dataset consisting of more than 12,000 secondary task engagements. Our main contributions are:

- The approach is more accurate than related work, predicting secondary task engagements in which long glances occur with 68 % accuracy and the total glance duration with a mean error of 2.4 s.
- The explanations replicate the results of various empirical studies and provide fast and easily accessible insights into the effect of touchscreen interactions, driving automation, and vehicle speed on driver distraction.

1.4 Outline

The three questions formulated in Section 1.3 form the leitmotif of the three main parts of this thesis as illustrated in Figure 1.2.

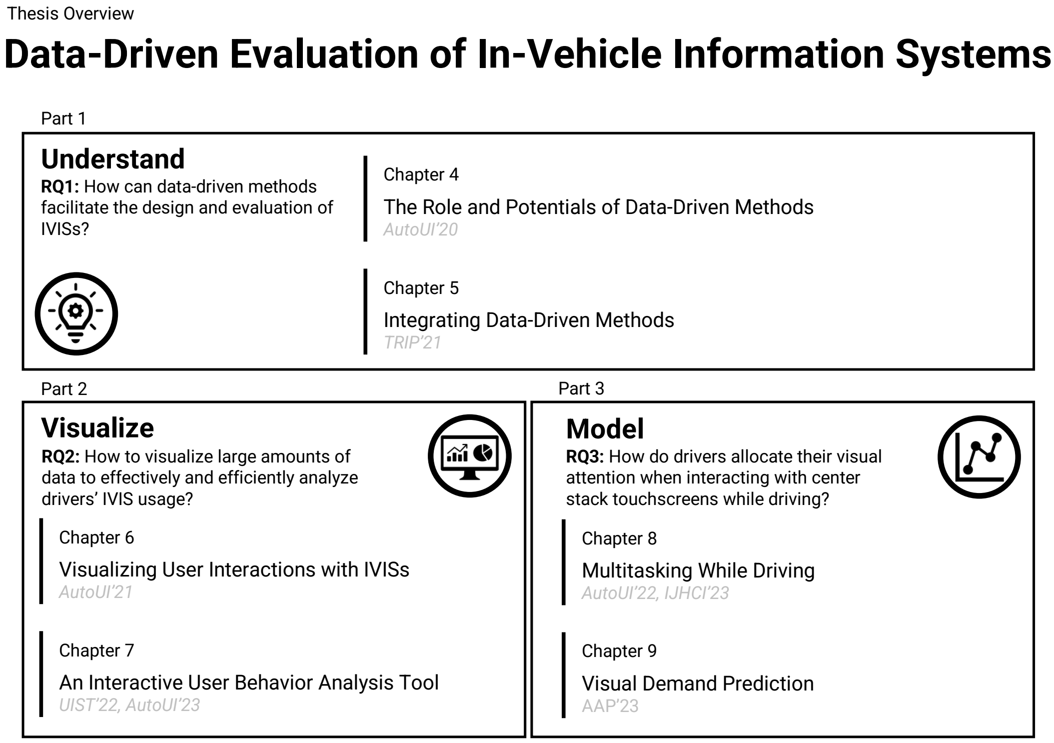


Figure 1.2: Visual abstract highlighting the three main parts of this thesis. Each part is represented by a block that corresponds to the chapters of this thesis and addresses a specific research question.

To put the scientific contribution of this thesis in context, we embed the three main parts in the remainder of this thesis. Chapter 2 lays the foundation necessary to understand the following chapters. We introduce the terminology used throughout this thesis, provide a general understanding of the theoretical construct of UX, introduce driver distraction, and explain why UCD is key to designing IVISs that are safe to use and meet user needs. We then introduce driving as a multitasking problem and explain why driver distraction is a safety risk that must be mitigated. In particular, we discuss the visual demand of in-vehicle touchscreen interactions and how to predict and explain drivers' visual attention allocation. In Chapter 3, we present the Telematics Data Logging Framework and introduce the data collection and processing procedures that form the basis for the subsequent visualization and modeling approaches. Finally, Chapter 10 concludes the thesis, revisits its contributions, and provides future research directions.

Throughout this thesis we use company names such as *Mercedes-Benz* or *Tesla* and product names such as *PySpark* or *AndroidAuto*. Please note that these are trademarks and we use them as such.

Chapter 2

Background and Related Work

The design and evaluation of touchscreen-based IVISs is an important factor in the automotive product design process. However, customer needs for a high level of functionality and efficient interaction with IVISs [10] often conflict with safety regulations and guidelines [13, 15, 39], as interactions with IVISs while driving pose potential safety risks. Thus, one of the main challenges facing designers and automotive OEMs is how to maximize the benefits of IVISs while keeping the risks associated with their distraction potential within acceptable limits [40]. In this chapter, we introduce the theoretical background and present related work necessary to understand how data-driven approaches can help address this trade-off. To understand how data-driven methods can support the design of automotive interfaces that are enjoyable and safe to use, a common understanding of UX and its role in the automotive industry is necessary. In Section 2.1, we present the basics of the UX concept and discuss its practical implementation in the design process. We then discuss how data-driven decision-making can improve product design and introduce different types of data that are used throughout the UX design process. Next, we present visual analytics and models of human behavior as data-driven solutions to improve the UX and safety of IVISs (see Section 2.2). Finally, in Section 2.3 we introduce driving as a multitasking problem, define *driver distraction*, and introduce the concepts of *visual demand*, *self-regulation*, and how they can be modeled. This chapter is partly based on previous publications [1, 2, 3, 4, 5, 6, 7, 8].

2.1 User Experience and Its Role in the Automotive Industry

Despite its long history in Human-Computer Interaction (HCI) and several attempts to define it [41, 42, 43], the term User Experience (UX) is still associated with a variety of meanings and used for different concepts. Then, in Section 2.1.1, we establish a common understanding of *User Experience* and its relationship to *Usability*. Then in Section 2.1.2 we discuss the differences in the understanding of UX in research and industry and what this means for the design and evaluation of IVIS. In Section 2.1.3 we introduce the UCD process and map its activities to the different phases of the design process. Finally, we discuss the role of UCD in practice (Section 2.1.4).

2.1.1 User Experience and Usability

User Experience is a holistic but fuzzy concept [44] for which a universal definition is hard to find [41]. Although there is a consensus that UX goes beyond the notion of usability [45], both terms are often used interchangeably and researchers and practitioners alike find it difficult to articulate the differences [46].

To create a common understanding of usability and user experience we use the notion of *do-goals* (e.g., to buy a new bike) and *be-goals* (e.g., to be proud of riding a very fast road bike) introduced by Hassenzahl [47]. Hassenzahl [47] argues that the experience of people interacting with a product can be divided into *pragmatic qualities* and *hedonic qualities* and that both together form the experience when interacting with products. *Pragmatic quality "refers to the product's perceived ability to support the achievement of 'do-goals'"* [47]. Thus, it focuses on the product itself, its utility, and usability with respect to an associated task. However, according to Brooke [48] *"[u]sability does not exist in any absolute sense; it can only be defined with reference to particular contexts"*. These points are well captured in the definition given in ISO 9241-11 where usability is defined as the *"extent to which a system, product or service can be used by specified users to achieve specified goals with effectiveness, efficiency, and satisfaction in a specified context of use"* [49]. Therefore, in the light of this work, we define the usability of an IVIS as the *extent to which the IVIS can be used by the driver to perform a specific task with effectiveness, efficiency, and satisfaction while driving a car.*

However, the quality of interactive technology goes beyond effectiveness and efficiency. [46]. In contrast to the pragmatic quality of an interaction, which is described by its utility and usability, the hedonic quality *"refers to the product's perceived ability to support the achievement of 'be-goals'"* [47]. Thus, the hedonic quality of an experience goes beyond the instrumental and encompasses more general human needs such as the need for social relatedness [46] or self-expression [47]. Hassenzahl [47] further argues that the fulfillment of be-goals is the main component of user experience. Consequently, ISO 9241-210 defines UX as *"a person's perceptions and responses that result from the use or anticipated use of a product, system, or service"* [50].

Thus, although pragmatic quality, i.e. usability, is not itself the primary and sole goal of UX design, it still facilitates the potential fulfillment of be-goals [47]. In other words, a lack of usability can be a barrier to good UX because problems cause negative emotions and negative emotions lead to bad experiences [46]. This makes usability an important part of a product's UX [51].

2.1.2 User Experience and Usability in Academia and Industry

The importance of UX is widely recognized in the automotive industry, and good a UX is the primary goal of most product development processes. Although the term UX is ubiquitous in academia and industry, the understanding of UX in these two areas is different [45]. While academic UX research focuses on theories, models, and frameworks, the main goal of industrial UX design is product improvement in terms of functionality, usability and novelty. According to Law [52] *"UX as a practice is an ongoing negotiation between researchers and practitioners"*. However, the question that remains is how to evaluate the UX to ensure that we are improving the product design based on the users' needs. Due to the intangible nature of UX, there is not one method, but rather a repertoire of methods and measures that can be used to assess the various dimensions (e.g., usability) of UX [52]. This leads to challenges in practice, as practitioners expect results to be presented in a concise and actionable manner [52]. Furthermore, the product design process is fast-paced and limited resources are available for UX evaluations, which should be performed continuously and as early as possible in the process [45].

While academic research seems to have already gone beyond the concept of usability, the question of how to design usable interfaces and how to evaluate usability is still a major issue in industrial applications. This is particularly true for the design of IVIS, and becomes evident when one considers that the usability of IVIS has been the biggest source of usability problems for new car owners for several years [35, 36, 37]. However, several domain-specific challenges apply to the automotive UX design process, and in particular to the design and evaluation of IVISs that need to be addressed and complicate the design task.

On the one hand, usability is strongly influenced by the context in which the interaction takes place [10, 38]. The experience of interacting with IVISs depends on environmental conditions (e.g., the driving scenario), the dual-task environment, and the frequency of use [53]. Thus, in addition to the physical and graphical interface design, designers must also address the influence of the driving situation [54]. This context dependency further increases the complexity of the design task [55].

2.1.3 The User-Centered Design Process

Given the current shortcomings of IVISs and the complexity of the design task, User-Centered Design (UCD) is considered the key to creating usable interfaces that are safe to use while driving. Accordingly, the automotive industry is moving from a technology-driven development approach to a more user-centered one [31]. UCD is based on four main principles (ISO 9241-210:2019 [50]): (1) Focus on the user (Who will use the system? In what context will they use it?), (2) Determine the design aspects that are important to users in the real application of the product, (3) Design iteratively and with continuous validation, (4) Consider the product as a complete system.

To improve the design and evaluation of IVISs, we focus on tools and methods that help UX experts to better understand how users interact with IVISs, which problems they are facing, and how the interface design affects their driving and interaction behavior. To understand how data-driven methods can help UX experts with these tasks, we need to create a common understanding of the UCD process and its associated methods. In line with Chen and Terken [56] we divide the UCD process into 6 stages: (1) *Strategy and Planning*, (2) *Knowledge Generation*, (3) *Design Generation*, (4) *Realization*, (5) *Evaluation*, followed by the (6) *Evaluation* process for further product improvements (see Figure 2.1). The individual phases can be directly mapped to the *Pre-Design*, *Design*, and *Post-Design* phases introduced by Nielsen [57]. In the following, we elaborate on the three design phases according to Nielsen [57] and the associated UCD activities in more detail.

The *Pre-Design Phase* is critical to ensuring that the subsequent process is well-informed and user-centered. It is composed of two main steps: *Strategy and Planning* and *Knowledge Generation*. The strategy and planning step defines the design rationale and determines various organizational aspects. The knowledge generation step has two main objectives: understanding the target user population and understanding the user tasks [57]. Various methods can be used to extract design-relevant knowledge about user characteristics, user tasks, and the context of use. They range from market research to questionnaires and interviews with current or potential users, to observational studies that identify usability bottlenecks by observing how users interact with the existing system [57]. Regardless of the methodology, the output of the pre-design phase is a set of functional and non-functional requirements that describe what the system should do and how it should perform (e.g., usability criteria) [56]. Overall, the pre-design phase provides the foundation for a successful design process by helping to establish a clear understanding of the users, their needs, and the context in which the product will be used.

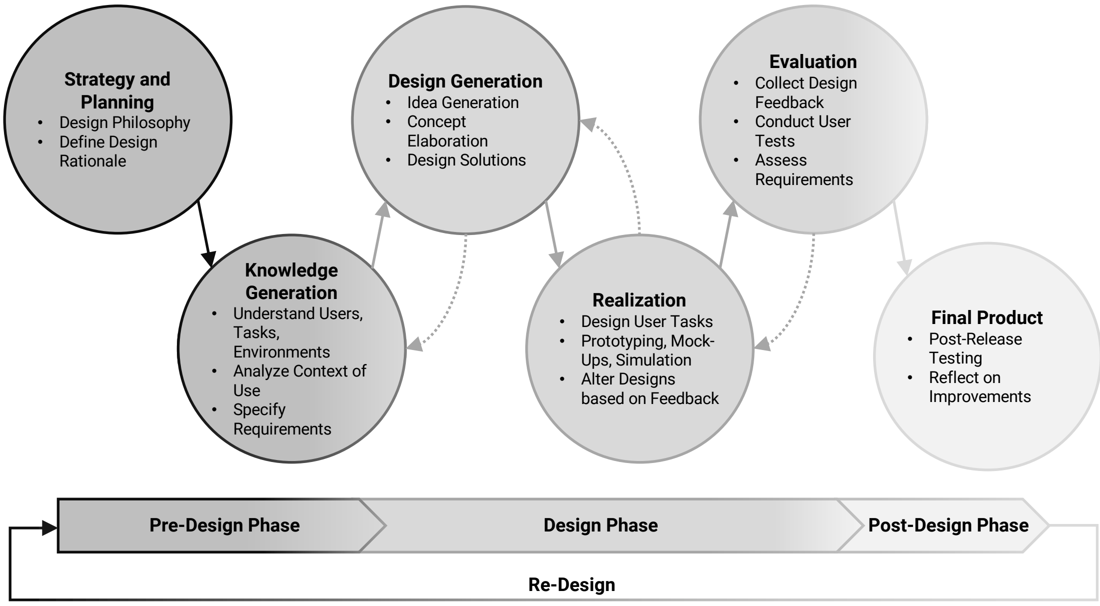


Figure 2.1: The UCD process, its activities, and associate design phases. The dotted lines represent re-iterations.

The main objective of the *Design Phase* is to derive a usable implementation that can be released and represents the final product [57]. The design phase consists of three iterative steps, namely *Design Generation*, *Realization*, and *Evaluation*. Due to the iterative design approach, which gradually refines the user interface based on continuous user feedback, these steps do not follow a strict order and are often repeated several times. As a result, the boundaries between the steps within the design phase are often blurred and tend to overlap. During design generation, initial ideas are translated into design concepts. User tasks and user-system interactions are designed and early concrete solutions, manifested in sketches or wireframes, are generated [50]. Based on these design concepts, the early design solution is made more concrete in the realization step. Here, designers create initial prototypes, mock-ups, or simulations. These tangible solutions can then be assessed using usability methods and empirically evaluated with real users in the evaluation phase. The main goal of this phase is to verify that the design solution meets the requirements.

As shown by the dotted lines in Figure 2.1, the central 4 phases of the UCD process are iterated until the requirements are met. This also indicates that the maturity of the product increases with each iteration, moving from an initial prototype to a usable solution that can be released. As a result, the activities in each phase change as the product matures. For example, in the early phases, low-fidelity click dummies may be qualitatively evaluated with internal participants in a think-aloud experiment, while in later phases, pre-production versions of IVISs may be quantitatively evaluated in driving simulators for potential driving safety risks.

The *Post-Design Phase*, begins after the product is launched and is particularly important in the redesign process of a system. A good understanding of how customers interact with the system in real-world conditions is essential to derive suggestions for improvements that can be implemented in a re-design. In addition, once the product is released, customer data can be used to evaluate initial usability goals and identify unintended usage patterns. User feedback can be collected either explicitly through methods such as customer surveys, or implicitly through data collection mechanisms that record user behavior at runtime. The latter allows companies to collect a vast amount of usage data that can be used to further improve product design.

2.1.4 User Centered Design in Practice

One of the main principles of UCD is to evaluate designs as early and as often as possible in the design process. Ideally, this is done by testing the design solution with real users and collecting their feedback. However, user-centered design and user-based testing are expensive. Based on a large survey conducted among UCD practitioners Vredenburg et al. [58] state that the costs associated with UCD methods often exceed 10 % of the total project budget. While 72 % of the respondents also reported that the integration of UCD methods significantly benefited the product development process within their respective organizations, the results of Vredenburg et al. also show that the choice of UCD is subject to strong cost-benefit considerations. Although field studies and user requirement analysis were considered most beneficial by practitioners, they were rarely used, and practitioners tended to choose informal, low-cost methods such as heuristic evaluation or informal expert interviews [58]. While the benefits of user-centered design methods are undisputed, their application, especially in early design phases, is not always practical [50]. In these cases, evaluations based on user simulation or computational models that predict user behavior are a promising alternative [59]. Although users are not directly involved, methods built on customer usage data or evaluated with data obtained from users are still user-centered [50] and can effectively complement traditional approaches.

2.2 Data-Driven Methods to Improve the Automotive UX Design Process

The continuous collection and analysis of customer usage data is already standard in web and app development and is used to support decision-making [60, 61, 62] and continuous product improvement [63, 64]. This is different in the automotive domain, where the decision-making culture, technology, and organization have been slow to adapt [65]. Despite the anticipated improvements, the area of data-driven methods that use large amounts of usage data obtained from customer vehicles to improve the design and evaluation of IVIS is not well explored. For example, Orlovska et al. [66] show that although modern cars collect large amounts of driving and interaction data, UX experts still lack methods and tools to use the data for IVIS evaluation. As a result, the problem that traditional methods are too slow and do not scale to the complexity of the problem is only shifted, not solved. In the following, we introduce the concept of data-driven decision-making (Section 2.2.1), discuss the specific characteristics of automotive data (Section 2.2.2) and present visual analytics (Section 2.2.3) and user modeling (Section 2.2.4) as potential approaches to improve the automotive UX design process and in particular the evaluation of IVISs.

2.2.1 Data-Driven Decision-Making

Data-driven decision-making can be described as *"the practice of basing decisions on the analysis of data rather than purely on intuition"* [67]. For example, a designer might simply choose one of two design alternatives based on experience and gut feeling. Alternatively, they could base their decision on the prediction of a user model derived from customer usage data that indicates how customers are likely to respond to the two alternatives. Or they could combine both approaches to arrive at a solution. This example also illustrates that data-driven decision-making is not always about making a clear distinction between data-driven approaches and intuitive or experience-based approaches [67]. Thus, data-driven decision-making can range from informing decisions with objective insights derived from data to large-scale automatic decision-making. Research shows that the adoption of data-driven decision-making is correlated with an increase in productivity. Brynjolfsson et al. [68], for example, show that the more a company relies on data in its decision-making processes, the higher its productivity. Considering that a large portion of the market value of today's top 100 digital companies is due to the data assets they create and use [67], the importance of data-driven decision-making and usage data collection becomes evident.

Following the example of digital companies, automotive companies also realized that the usage of big data can be a competitive advantage [65]. Concerning the design of IVISs, data can play an important role throughout all phases of the UX design process (see Figure 2.1). To make decisions that are centered around the user, it is inevitable to create a holistic understanding of who the users are, how they behave, and what they desire. While it is difficult, if not impossible, to measure the fuzzy construct of UX itself, it is possible to measure and improve certain aspects of it (e.g., usability). Therefore, our goal is not to provide a data-driven approach to measure UX or to automate the decision-making process, but to develop data-driven methods that facilitate various UX activities within the automotive UX design process. We want to enable practitioners to make better and more user-centered design decisions by incorporating insights derived from data.

2.2.2 The Characteristics of Automotive Data

A variety of data can inform decision-making and can be used as the basis for data-driven methods that inform decision-making. In general, this data can be divided based on its type (*Qualitative Data* vs. *Quantitative Data*), the way it is collected (*Implicit Feedback* vs. *Explicit Feedback*), and the environment and context in which it is collected (*Lab Data* vs. *Naturalistic Data* vs. *Natural Data*). To create a common understanding and to better understand the focus of this thesis we elaborate on the different characteristics.

Qualitative Data vs. Quantitative Data Qualitative research methods focus on the quality of things and are used to explain, describe, and identify the cause of user behavior [69, 70]. Deniz and Lincoln [71] describe qualitative research methods as an approach to interpret phenomena in their natural environment based on the meaning that a particular user or group of users reveals to them. Therefore, qualitative methods usually focus on collecting subjective impressions of participants that they give to behaviors, events, or objects. Qualitative research focuses on explaining why certain behaviors or phenomena occur, how certain aspects of design are perceived, and why people behave in certain ways.

Typical qualitative data collection methods include interviews, focus group discussions, or observations [72]. These methods can be used in different stages of the design process.

Quantitative research, on the other hand, focuses on measurements to test hypotheses, determine and quantify an outcome, identify correlations, and generalize findings [71]. Thus, the main goal of quantitative research is to explain a particular phenomenon by collecting numerical data and analyzing it using statistical methods [69]. When quantitative methods are used properly, hypotheses can be tested for statistical significance and statements can be made with some confidence that the result is not just an effect of random noise. The difference between qualitative and quantitative data can be broken down into the types of questions that can be answered: While quantitative approaches aim to answer questions such as *"How many?"* and *"How much?"*, qualitative approaches aim to answer the question of *"Why?"* something happened or a *"Why?"* a participant behaved in a certain way.

Explicit Feedback vs. Implicit Feedback The data used in the product development process can also be categorized considering the method of collection. There are two different types of feedback used to evaluate user behavior, explicit and implicit feedback. *Explicit Feedback* is collected intentionally, for example, through surveys, focus groups, or interviews. People deliberately provide the information they were asked for. On the other hand, *Implicit Feedback* is gathered through observation and recording of user interactions with technology, rather than being explicitly provided. This type of feedback can be obtained through sensors, such as cameras and speed sensors, or by logging user interactions with the product. This work only focuses on implicit feedback collected from vehicle actuators, sensors, vehicle apps, or in-vehicle software systems. Therefore, the explicit quantitative feedback generated by extensive user surveys, or by using an automated data collection method, such as web surveys, will not be considered in this work.

Lab Data vs. Naturalistic Data vs. Natural Data To evaluate driver behavior and driver interactions with IVISs it is important to consider the context in which the data is collected. Implicit driver behavior data can be collected in laboratory experiments, during naturalistic driving studies, or directly from customers. We, therefore, differentiate between *Lab data*, *Naturalistic Data*, and *Natural Data*.

Lab data is data collected during controlled experiments in an artificial environment. For most lab experiments a small number of participants are recruited and instructed to perform specific tasks in a driving simulator environment (see [73, 74, 75, 76]). The fidelity of the driving simulator can range from a simple seating buck [75] to a high-fidelity moving base driving simulator [76]. The experimental setting makes it easy to supplement the implicit feedback collected during the experiment with qualitative and explicit feedback (e.g., by conducting follow-up interviews). However, a large investment of time and resources is required, and the number of participants is strictly limited by the available budget.

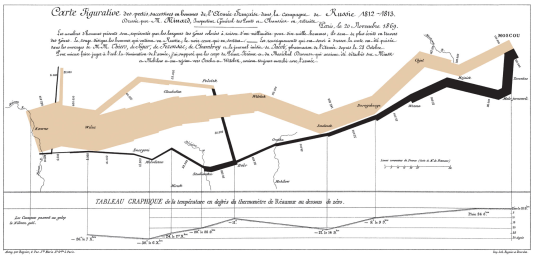
Whereas lab data is collected in an artificial environment, *Naturalistic Data* is gathered during driving *"in a natural driving context and under various driving conditions"* [38] rather than in a controlled laboratory environment. The goal of naturalistic driving studies is the unobtrusive collection of data from participants to study their behavior in real-world conditions. Conducting naturalistic driving studies requires a considerable amount of resources. A representative sample of participants who are willing to contribute with their data over several weeks to months must be recruited, participants' vehicles must be equipped with the necessary sensors, and data collection, processing, and analysis pipelines must be developed [23, 77, 78, 79]. Similar to studies conducted in a controlled laboratory environment, follow-up interviews or questionnaires aimed at triangulating quantitative findings with qualitative ones may be part of naturalistic driving studies. Prominent examples of large naturalistic driving studies are the Strategic Highway Research Program 2 (SHRP2) [79], the 100-Car Study [78], the MIT Advanced Vehicle Technology Study [77], and the UDRIVE study [80].

In contrast to naturalistic data, *Natural Data* describes data collected from customers who provide their data by purchasing the vehicle and agreeing to the terms of data collection. While data is collected under real-world driving conditions similar to naturalistic driving studies, customers are not explicitly participating in a study but rather providing their data to the OEM for product improvement. This implies that natural data is collected using only the existing and available means of the original production vehicle. This allows natural data to be collected from any vehicle in an OEM's fleet over an undefined period. However, to collect natural data from a customer's vehicle, privacy regulations such as the General Data Protection Regulation (GDPR), which restrict the collection of personal data, must be met.

Each type of data has its advantages and disadvantages. When comparing lab data with naturalistic and natural data, it is important to note that the reliability of the derived results depends not only on the data collection environment, but also on the quality of the sensors, the representativeness of the sample, and the amount of data available. Laboratory experiments and naturalistic driving studies allow the extraction of very specific and detailed information, either through additional qualitative feedback or through sensors that are not available in production vehicles. However, conducting these studies is expensive and time-consuming which limits the amount of data that can be collected. Natural data on the other hand can be collected automatically and in large quantities. This opens up many areas of application, some of which will be presented in the remainder of this thesis.

2.2.3 Visual Analytics

In today's product development, decision-makers, regardless of the domain, want to enhance their decisions with insights from customer data. This allows them to design products that are tailored to customer needs [81]. However, the main challenge in data-driven decision-making is not the acquisition of raw data itself, it is the challenge of extracting useful knowledge from it [82]. To create solutions that help designers, engineers, or scientists, the right information must be available at the right time [83]. Therefore, tools and analytical solutions need to communicate the results of analysis through meaningful visualizations and clear representations [83]. The latter describes the overarching goal of visual analytics research. Accordingly, Keim et al. [84] provide the following definition: *"Visual analytics combines automated analysis techniques with interactive visualizations for an effective understanding, reasoning and decision making on the basis of very large and complex data sets."*



(a) Map by Charles Minard from 1869 showing the loss of soldiers, troop movements, and temperatures over the course of Napoleon's Russian campaign. [87]

(b) OutFlow as proposed by Wongsuphasawat and Gotz [88]. In contrast to the original Sankey diagram it incorporates temporal information as visualized by an additional edge between two nodes.

Figure 2.2: Visualization examples.

However, creating meaningful visualizations and intuitive tools for analyzing big data is far from obvious. Although several commercial general-purpose tools exist, they often fail to meet domain and task-specific needs. Domain experts often require advanced visualization and interaction concepts that are not supported by commercially available tools. These tools stick to a small set of standardized visualizations [85].

While these standardized dashboards are a valuable tool for quickly communicating Key Performance Indicators (KPIs) to stakeholders or managers, they are often disconnected from the domain expert's workflow and serve as a reporting tool rather than an exploratory knowledge generation tool. To support domain experts in their work, it is important to create solutions that are specific to their workflow. Individual visualizations often address a specific task that is part of a larger workflow. Therefore, these visualizations need to be linked in such a way that they support this workflow as a whole [86]. An example of this is the common need to explore large amounts of data at multiple scales. This can be achieved by visualizing the data on different levels of granularity, starting broadly and zooming in on details as the analysis progresses [81]. In addition, domain experts are often non-specialists when it comes to analyzing large amounts of data. Thus, it is important to avoid information overload and that visualizations are easy to understand and their benefits immediately apparent Keim et al. [83].

Visualizing Event Sequence Data There is not much research on visual analytics to evaluate driver interactions with IVISs. Most approaches in the automotive domain focus on visualizing data collected from a few sensors [89] or in controlled experiments [90]. For example, Jansen et al. [90] present an approach to visualize spatiotemporal data collected during user interface interaction studies. Their approach provides valuable insights into the combined visualization of explicit and implicit feedback collected from different sources. However, the tool is designed for academic needs and focuses on the visualization of individual situations recorded during user studies. Therefore, it does not apply to the challenges faced by industrial UX experts when they need to analyze millions of data points collected from customers.

However, outside the automotive domain, there is a lot of research on the analysis of so-called *Event Sequence Data* [62]. Event sequence data describes data that consists of multiple series of time-stamped events ordered over time. It includes website logs describing how users navigate pages [91, 92, 93], medical event data [94], or car service records [95]. Regardless of the particular use case, the main application is to compare and analyze different sequences of events (e.g., *Homescreen → Settings → Privacy and Security*), based on, for example, their frequency (e.g., 35% of users went from *Settings* to *Privacy and Security*), the time intervals between events (e.g., it took them 5 seconds on average), or the number of glances drivers needed to perform a sequence of interactions.

Sankey Diagrams [96, 97] are a well-known and widely used group of event sequence visualizations (see Figure 2.2a). Sankey diagrams are directed graphs consisting of nodes and links that visualize quantitative information about flows, their dependencies, and their partitioning into different paths. Each node represents a state in a flow and has weighted input and output links (except for source and sink nodes). Links represent transitions from a source node to a destination node. The weight of the links represents the flow quantity, visualized as the width of the link. Except for source and sink nodes, the sum of incoming links is equal to the sum of outgoing links. While they can effectively visualize flows between different nodes, Sankey diagrams, originally, do not take into account the temporal aspect of transitions. An approach that addresses the processing of temporal event data is presented by Wongsuphasawat et al. [98] and is called *LifeFlow*. The approach combines multiple event sequences into a tree while preserving the temporal spacing of events within a sequence. Whereas in LifeFlow multiple event sequences are combined in a tree, *OutFlow* [88] combines them into graphs, similar to Sankey diagrams (see Figure 2.2b). To represent the temporal spacing between events, the authors introduce an additional type of edge whose width represents the duration of the transition. Sankey diagrams, LifeFlow, and Outflow, all focus on visualizing and analyzing the different flows, their distribution, and their temporal aspects from one dataset. In contrast, the *MatrixWave* approach presented by Zhao et al. [93] aims to create a comparative analysis of multiple event sequence datasets by replacing the edge connector of the Sankey diagrams with transition matrices. Whereas the aforementioned approaches are solely focusing on visualizing event sequence data, other approaches aim to provide an overall framework for user behavior evaluation in a digital environment [99]. In addition, commercial providers like UserTesting[1], UserZoom[2] and alike offer tools to analyze user sequences. However, to meet the requirements of automotive UX experts, an approach has to be developed that allows to analyze event sequences on the one hand and provides direct insights into driving behavior and gaze behavior on the other hand.

2.2.4 Modeling of Human Behavior

Visual analytics can provide a lightweight entry point to big data analytics by allowing users to explore and interact with large amounts of data. Users can identify trends, compare different distributions, and generate insights. However, the associated visualizations not only support insight generation but can also be used to present and communicate analytical results in the appropriate context [100]. Thus, they can effectively support analytical reasoning and decision-making [101]. However, when it comes to quantifying

[1] https://www.usertesting.com
[2] https://www.userzoom.com

effects, exploring previously unknown and complex relationships, or making predictions about human behavior, data visualizations reach their limits. Here, sophisticated statistical and computational models can further support data-driven decision-making by testing design hypotheses for significance or modeling user behavior so that designs can be evaluated without the need for human participants. We distinguish between statistical models, which generate knowledge about human behavior and computational models, which aim to predict and simulate human behavior. This distinction borrows from the distinction between explanatory and predictive modeling made by Shmueli [102]. While both types of modeling can be used for either purpose, in HCI statistical models are primarily used for explanation and computational models for prediction. This distinction will be discussed in more detail below.

Statistical Models to Generate Knowledge About Human Behavior Problems in HCI are often concerned with the impact of technology on human behavior. Oulasvirta and Hornbæk [103] describe HCI as problem-solving and state that the overarching goal of HCI research is to develop the *"ability to solve important problems related to human use of computing"*. This implies that in order to assess this ability, one needs to measure the effect that a proposed solution has on problem-solving ability [103]. Other research questions in HCI are concerned with the general effect of technology on human behavior or the effect that specific changes in the agent, environment, or scenario have on the interaction with technology [104].

Regardless of the problem, these effects are often evaluated in an experimental setting, where the effects of independent variables on dependent variables are measured in a controlled environment to either accept or reject a research hypothesis [105]. However, because researchers do not have access to the entire population, statistical models come into play. These models allow generalization from a sample to the population from which the sample was drawn [106]. This generalization is called statistical inference. However, there are two different branches of statistical inference: Frequentist and Bayesian statistics. The main difference between these two types is that in Bayesian models, latent variables such as model parameters are described as probability distributions, whereas in frequentist models they are fixed. Thus, whereas frequentist models provide a point estimate (p-value) that indicates whether a hypothesis shall be accepted or rejected, Bayesian models provide a probability for or against a given hypothesis. Thus, Bayesian models can be used to quantify uncertainty where frequentist models provide only a yes/no answer.

Nevertheless, the rigorous framework for hypothesis testing, the well-established methodology, and the reproducability makes frequentist approaches well-suited for various applications within HCI. This also applies for the evaluation of driver behavior and human-vehicle interaction. Grahn and Kujala [107], for example, use multilevel modeling [108] to evaluate the effect of touchscreen size and user interface design on the visual demand of touchscreen tasks and driver distraction. Another example is provided by [27], who use a logistic regression model to model drivers' tactical self-regulation.

Computational Models to Predict and Simulate Human Behavior Whereas statistical models are often used for explanatory modeling [102] to generate insights given a particular set of data, computational models are formal representations of systems or processes, expressed in code or mathematics, that aim to replicate reality to make predictions [59].

Accordingly, Wilson and Collins [109] state that the *"goal of computational modeling in behavioral science is to use precise mathematical models to make better sense of behavioral data."* These models can also be used to evaluate the expected value of a design, as well as to predict, explain, and even shape user behavior [59].

In line with Banovic et al. [110], we divide computational models into theory-driven or theoretical models and data-driven or algorithmic models. Theory-driven computational models are based on theoretical assumptions about human-technology interaction. These may be theories of perception, cognition, or motor control. While it is tempting to use sound theories as a basis for modeling and then fit these models to specific use cases to make predictions about human behavior, theory-driven models often lack predictive power [102] and practical application [111].

Data-driven computational models, on the other hand, use data mining and machine learning techniques to extract patterns from large amounts of data. Based on these patterns, they can make predictions about future behavior. A key difference between statistical and data-driven computational models is how they treat the underlying data. While statistical models assume a stochastic data model, data-driven computational models make no assumptions about the distribution of the data. They aim to find a function or algorithm that operates on the input variables to predict the target variables [111]. Using the aforementioned data mining and machine learning techniques, data-driven computational models have shown promising results in various areas, such as making design suggestions for user interfaces [112, 113, 114, 115], learning human mobility patterns [116, 117], and predicting human-vehicle interactions [118, 119, 120].

We argue that the use of large amounts of driving and interaction data in combination with data-driven computational models and in particular machine learning approaches can be a promising step towards more accurate and holistically applicable tools to support the evaluation of IVISs. However, two main factors prevent these approaches from providing valuable insights into the interaction between driver and IVIS. First, interaction data is not yet available in the same quantity as driving data. Second, most of the above-introduced supervised machine learning approaches lack explainability. While they report use-case-specific performance metrics, the models remain a black box, providing no insight into the features that are decisive for the predictions.

2.3 Driving: A Multitasking Problem

Driver distraction poses a significant threat to road safety [40]. The National Highway Safety Traffic Association (NHTSA) [14] reported that in 2020, 3,142 fatalities and 424,000 injuries resulted from crashes caused by distracted driving. Despite the increasing availability of driver assistance and monitoring systems, the number of accidents related to distracted driving has not decreased. In the following, we define what driver distraction is, how it affects road safety, and elaborate on how drivers divide their visual attention when interacting with IVISs.

2.3.1 What is Driver Distraction?

Driving is a complex multitasking activity that requires sustained attention and resources. Drivers need to perform several activities simultaneously [40]. They need to watch and follow the road, perform steering and pedal movements, and react to sudden changes in the driving environment [121]. Despite the complexity of the primary driving task, drivers tend to engage in non-driving related secondary tasks like talking to the passenger or interacting with the smartphone or the IVIS. However, as human attention is limited, these interactions compete with the resources required to perform activities critical for safe driving. Consequently, Lee et al. [122] describe the interaction with devices like mobile phones or IVISs as competing activities. The explicit attention toward a competing activity is also what differentiates *distraction* within the broader scope of *inattention*. Whereas inattention can also occur due to a cognitive state (e.g., fatigue or drowsiness), distraction always involves an explicit activity that leads to a diversion of attention away from driving [122]. Thus, in alignment with Lee et al. [122], the following definition of driver distraction guides the remainder of this thesis:

> *"Driver Distraction is a diversion of attention away from activities critical for safe driving toward a competing activity"* [122].

However, Kircher and Ahlstrom [123] point out that although the term *"safe driving"* considers the influence of the driving context (i.e., driving demand), it is not sufficiently defined. They argue that without a clearly defined baseline for what is considered safe driving, current definitions suffer from hindsight bias and are difficult to operationalize. Accordingly, Kircher and Ahlstrom [123] introduce the theoretical concept of *minimum required driver attention*, which extends current definitions by defining what a driver must pay attention to. They also consider the drivers' ability to self-regulate demands and create spare capacity.

Despite the slight differences in the various definitions that exist, driver distraction can be further classified according to the sensory modality of the distraction [124]. Accordingly, Regan et al. [124] distinguishes six types of distraction, namely *visual distraction, auditory distraction, cognitive distraction, olfactory distraction, gustatory distraction, and tactile distraction*. However, to evaluate driver interactions with the center stack IVIS, we will focus on visual distraction.

Visual distraction can be described as *"[d]iversion of attention to things we see"* [121] or as *"taking one's eyes off the road"* [125]. However, according to Regan and Oviedo-Trespalacios [121] the latter rather describes a response that can be triggered by various types of distraction. For example, an alerting sound (auditory distraction) may cause the driver not only to take his ears off the road but also to take his eyes and hands off the road to answer an incoming phone call by clicking on the touch screens. The responses caused by distractions thus interfere with the activities critical for safe driving and are therefore considered a road safety risk.

2.3.2 The Effect of Driver Distraction on Driving Performance and Road Safety

Driver distraction is a significant factor in road safety, as observed in naturalistic driving studies and crash data analysis [79, 126]. Analyzing data of the SHRP2 naturalistic driving study, Dingus et al. [126] estimate that up to 68.3% of injurious and property damage crashes involve general distraction. They further state that potentially 36% of crashes could be avoided if distractions were eliminated [126].

Various studies [16, 79, 126, 127, 128] show that out of the many sources of distraction (e.g., talking to a passenger, daydreaming, listening to music), visually demanding tasks (e.g., interacting with the IVIS or the smartphone) that require the driver to take the eyes off the road are associated with the highest road safety risk. This is in line with findings from Gershon et al. [129] who report that drivers' time spent looking away from the road accounted for 41% of the crash risk associated with manual cell phone use, compared to 10% of the risk associated with reaching/handling objects while driving. According to Dingus et al. [126], the risk of such visually demanding activities is two times the overall distraction risk. One accommodating factor is that taking the eyes off the road significantly impairs driving performance and slows down response times up to 29% in real traffic [130]. This does not only apply to manual and partially automated driving but also to higher automation levels [19]. Recent research shows that takeover performance is significantly affected by the visual-cognitive load of the secondary task [131] and distraction in general [132].

Thus, taking the eyes off the road for longer periods of time is correlated with increased crash risk [16, 133]. Klauer et al. [16] found that glances off the road longer than two seconds increase the crash risk by two times compared to normal driving. Accordingly, the "Visual-Manual NHTSA Driver Distraction Guidelines for In-Vehicle Electronic Devices" [15] define upper bounds for glances longer than two seconds. This shows that the visual demand of secondary task engagements is an important factor that needs to be considered when designing IVISs.

2.3.3 Drivers' Self-Regulative Behavior

While interacting with touchscreen-based IVISs, drivers divide their visual attention between the primary driving task and the secondary touchscreen task. Research shows that drivers actively self-regulate their multitasking behavior to maintain safe driving. They adapt their level of engagement to mitigate the risks associated with the secondary task demands [134]. According to Rudin-Brown [134], this self-regulative behavior can be intentional or unintentional and it occurs at three distinct levels derived from Michon's driver task model [135]: strategic, tactical, and operational.

Strategic self-regulation describes driver decisions that are made on a timescale of minutes or more [134]. These decisions are often consistent throughout a trip. For example, some drivers report that they never engage in a secondary task in heavy traffic, in bad weather, or when driving at night [136]. Oviedo-Trespalacios et al. [27] modeled strategic self-regulation as the decision to pull over to perform a secondary task [27]. In this study, some drivers made a strategic decision not to engage in a secondary task while driving.

Tactical self-regulation refers to a driver's decision to engage in a secondary task according to the driving demand. Drivers make tactical decisions in the time frame of seconds [134] and continuously update them while driving. Many studies have examined drivers' engagement in mobile phone tasks while driving. The results show that drivers are less likely to engage in a visual manual phone task when driving demands are high (high speed, sharp turns, etc.) [27, 137, 138, 139, 140]. Tivesten and Dozza [137] show that drivers use information about the upcoming driving demand to decide whether or not to engage in a secondary task. Somewhat contrary results are presented by Horrey and Lesch [141]. The authors found that, although drivers were well aware of the demands of specific traffic situations, it had little influence on the decision to engage in the secondary task [141]. This is consistent with the findings of Carsten et al. [142], who show that drivers stopped engaging in easy tasks when the driving demand increased but continued to engage in more demanding secondary tasks [142]. Liang et al. [143] found that drivers avoided initiating a secondary task prior to an immediate transition to higher driving demands. However, drivers did not postpone their secondary task engagement when driving demand was already high [143]. Carsten et al. [142] and Liang et al. [143] argue that more work is needed to assess the factors that influence tactical self-regulation.

Operational self-regulation describes behavioral adaptions made by the driver while actively engaging in a secondary task. This implies that the driver has already made a strategic and tactical decision to engage in a secondary task. Operational self-regulation can be bidirectional. Research shows that drivers adjust their driving behavior in terms of vehicle speed, lane position, or time headway when engaged in a secondary task [26, 144, 145, 146, 147]. On the other hand, recent evidence shows that drivers also adjust their engagement in secondary tasks in response to variations in driving demand. Oviedo-Trespalacios et al. [27] found that drivers temporarily stopped the use of mobile phones to cope with varying driving demands [27]. Similarly, in a test track experiment, Liang et al. [143] show that drivers adjust their time-sharing behavior according to driving demands. In addition, Tivesten and Dozza [148] state that drivers allow for more distraction in less demanding situations. In a naturalistic driving study, drivers performed shorter off-road glances during turning when a lead vehicle was present and when they detected oncoming traffic [148]. Tivesten and Dozza [148] further state that drivers prioritize secondary tasks over monitoring the driving environment, especially in low-speed situations. Accordingly, Risteska et al. [22] show that drivers' off-path glances decrease in situations with higher visual difficulty.

2.3.4 The Effect of Driving Automation on Self-Regulation

Many studies have investigated the effect of partially automated driving (Level 1 and Level 2 according to SAE J3016 [30]) on drivers' secondary task engagement. As laid out in the following, the results suggest that more automation results in less driver engagement and, thus, a lower capability to assess the current driving situation correctly.

Lin et al. [149] investigate drivers' self-regulation in Level 2 driving according to the levels of situation awareness as proposed by Schömig and Metz [150]. On the control level, which corresponds to operational self-regulation, they found that drivers adapt their behavior according to the severity of the hazard. Whereas they pause their engagement in case of urgent hazards, they continue to engage with a more frequent task-switching behavior in case of less urgent hazards. In addition, many studies investigated how drivers allocate their visual attention during partially automated driving. Results from the Virginia Connected Corridors Level 2 naturalistic driving study [24] indicate that the use of Level 2 automation (i.e., ACC+LCA) led to drivers spending less time with their eyes on driving-related tasks. In accordance, Gaspar and Carney [151] found that with partial automation activated, drivers made longer individual off-road glances and had longer maximum total-eyes-off-road times. This finding is complemented by the results presented by Yang et al. [152] who also found that off-road glances were longer in automated driving conditions and additionally investigated the effect of different levels of distraction. They found that off-road glances were longer for highly distracting secondary tasks [152]. Noble et al. [153] assessed the effect of ACC, LCA, and ACC+LCA on drivers' glance behavior and secondary task engagement. They found that during ACC+LCA driving, drivers execute longer and more frequent glances away from the road [153].

They, however, did not find significant differences in the mean off-road glance duration nor in the tactical self-regulation when ACC+LCA was active. Another naturalistic driving study is presented by Morando et al. [23] who found a significant decrease in the percentage of time with eyes on the road center when using ACC+LCA. In a subsequent study, the authors investigated drivers' glance behavior during disengagements of Tesla's Autopilot in naturalistic highway driving [154]. Whereas they found that all off-road glances tended to be longer with Autopilot compared to manual driving, the difference was particularly large for glances down and toward the center stack. The mean glance duration increased by 0.3 seconds and the proportion of glances longer than 2 seconds increased by 425 % in Autopilot conditions compared to manual driving.

2.3.5 Visual Demand of Secondary Task Engagements

To interact with touchscreen-based IVISs and to capture the information that is visualized on the display, drivers need to take their eyes off the road. However, taking one's eyes off the road during secondary task engagements compromises driving performance and safety [16, 133, 155, 156, 157, 158]. This makes assessing the visual demands of secondary tasks a critical aspect of road safety.

Visual demand is defined as the *"degree or quantity of visual activity required to extract information from an object to perform a specific task"* [39]. ISO:15007:2020 [39] proposes several glance-based metrics to measure visual demand. Two of the most widely used metrics are the total glance duration and the mean glance duration. The total glance duration is the *"summation of all glance durations to an area of interest (or set of related areas of interest) during a condition task, subtask or sub-subtask"* [39]. The mean glance duration is the *"mean duration of all glance durations to an area of interest (or set of related areas of interest) during a condition task, subtask or sub-subtask)"* [39].

The visual demand of touchscreen interactions with IVISs can therefore be described in terms of how long and how often the driver has to take the eyes off the road to perform a specific task. Consequently, these metrics are often used as a proxy to measure the influence of secondary task engagements on road safety. However, according to Victor et al. [79] there is no single metric that can fully describe the relationship between glance behavior and risk, but rather a combination of metrics is necessary. Even though any single measure provides an incomplete assessment of distraction, empirically supported measures such as those presented above should be used for decision-making as early as possible in the design process [159].

2.3.6 Visual Demand Prediction

Various methods aim to predict visual-manual distraction while driving [22, 119, 120, 160, 161]. Most of them focus on driver distraction detection to warn the driver when a potentially dangerous situation is detected. These approaches are often based on naturalistic driving data and employ various machine learning methods. They utilize driving performance metrics (e.g., speed or steering wheel angle) [119, 120, 160], environmental data (e.g., traffic conditions) [22], or video data of the driver [161, 162]. While these approaches show promising results, they do not incorporate any information on how drivers interacted with secondary devices like mobile phones or IVISs. Therefore, they do not generate insights into the visual demand of specific UI elements or interactions.

However, various approaches exist that model the visual demand of IVISs based on user interactions with specific UI elements. They explain the effect specific interactions have on drivers' visual distraction. These approaches focus on the understanding of interaction behavior and aim to identify distracting features of IVISs. Their purpose is to inform designers and researchers in the early stages of the development process about potential implications of their design on driver distraction. In this work, we focus on the latter and provide an overview of the current state-of-the-art in this domain.

Most computational models that predict visual demand based on user interactions are theory-driven computational models derived from Keystroke-Level Model (KLM) [163, 164]. In such approaches, an entire task is decomposed into a sequence of specific primitive operators (e.g., pressing a button, or searching a list). The interaction duration for each operator is then determined empirically [165] and the total time on task prediction is equal to the sum of the individual interaction durations of the respective operators occurring in the task. The KLM was originally developed to predict processing times in computer-assisted office work, but several adaptations have been made to assess IVISs [165, 166, 167]. However, most of these approaches focus on task completion times rather than visual demand. Pettitt et al. [168] were the first to propose a KLM-based approach to predicting visual demand. They show a high correlation between predicted values and measures from an occlusion experiment in which the driver's vision is occasionally occluded to simulate distracted driving [169]. The first KLM-based method to directly predict visual demand is presented by Purucker et al. [170], who propose a task-specific KLM model. They argue that the use of fixed operators to model innovative and new hardware is limited. While their approach can only predict the total glance duration, Large et al. [171] propose a method that can additionally predict the number of glances and the mean glance duration. Their information-theoretic approach is based on the Hick-Hyman law for decision/search time and Fitt's law for pointing time.

While the presented KLM-based approaches achieve promising results, they all share several drawbacks. First, due to their cumulative and linear nature, the models are not well suited to model potential (non-linear) dependencies between different user interactions or driving situations. For example, the difference in visual demand between selecting an item from a list and tapping a button may be negligible at low speeds, but significant at higher speeds. In addition, the length of an interaction sequence combined with specific interactions may also influence visual demand in a non-linear and non-additive manner [170]. For example, if the driver presses two buttons that are close to each other, this is unlikely to result in a doubling of the total glance duration, as the driver could perform both interactions in one glance. Second, the model parameters of the presented approaches are derived from empirical tests with driving simulators of varying fidelity and with relatively small numbers of participants. This is likely to lead to predictions that are biased to the specific experimental setting, as also noted by Large et al. [172] and shown in a real-world driving experiment evaluating the applicability of Fitt's law [173]. Third, current approaches do not take into account the effect of different driving situations on the visual demand of secondary tasks. Research shows that drivers modulate their task engagement and visual attention based on driving demands [22] and the degree of assisted driving [151, 154, 171, 174], making it important to include such parameters.

A different approach is taken by Kujala and Salvucci [175], who propose a model based on the Adaptive Control of Thought—Rational (ACT-R) cognitive model architecture [176]. Their approach aims to represent the visual sampling strategy of drivers. They argue that drivers adjust their glances based on a time constraint that depends on the current driving situation. While the model can predict multiple facets of visual demand, only grid and list layouts are considered. Furthermore, the driving scenario is relatively simple and the evaluation shows significant drawbacks in prediction accuracy, especially concerning the detection of long glances.

2.4 Definitions and Terminology

At this point, we will repeat some concepts already in use and introduce new ones that will be used throughout the rest of the thesis.

Primary Driving Task: The primary driving task consists of all activities critical for safe driving [121]. This includes driving the vehicle and maintaining alertness to traffic and other potential hazards [177].

Secondary Task: Secondary tasks are non-driving related activities that interfere with activities critical to safe driving [121]. They describe a diversion of attention from the primary task of driving to a competing activity.

Driver Distraction: Driver Distraction is a diversion of attention away from activities critical for safe driving toward a competing activity [122].

Secondary Touchscreen Task: A secondary touchscreen task describes a non-driving related activity on the UI of a touchscreen-based IVIS located in the center stack of the vehicle. It is defined as an objective that a user must solve and consists of a defined start and end. The start and end of a task can further be defined by one or multiple conditions, being for example specific UI elements. A task can consist of multiple flows, meaning that the progression of how a user went from the start to the end is arbitrary.

User Flow: A user flow describes a linear series of interactions (e, p), where e represents the type of UI element and p the gesture type.

Touchscreen Interaction: A touchscreen interaction $i = (t, e, p, c)$ is composed of its timestamp t, UI element type e, gesture type p and coordinate pair $c = (x, y)$.

Interaction Sequence: An interaction sequence $I = (i_n)_{n=1}^{N}$ is a sequence of touchscreen interactions recorded during one trip, where i_n is a single touchscreen interaction performed by a user and N denotes the number of interactions of I. Within I, the duration between two successive interactions $t(i_{n+1}) - t(i_n)$ must be smaller than or equal to Δt_{max} such that $t(i_{n+1}) - t(i_n) \leq \Delta t_{max}$.

Glance Sequence: A glance sequence $G = (g_n)_{n=1}^{N}$ is a sequence of non-overlapping intervals of driver glances, where g_n is a single glance performed by a user and N denotes the number of glances of G. Each glance $g_n = (t^s, t^e, r)_n$ is composed of its start time t^s, end time t^e, and Area of Interest (AOI) r, describing where the driver looked at between t^s and t^e. For all glances of a but the first of a trip, the start time is equal to the end time of the preceding glance $t_s(g_n) = t_s(g_{n-1})$.

Driving Sequence: A driving sequence $D = (d_n)_{n=1}^{N}$ is a sequence of driving data observations, where d_n is a single observation and N denotes the number of observations of D. Each observation is defined as $d_n = (t, v, \theta, a_{\text{ACC}}, a_{\text{LCA}})_n$, where t represents the timestamp, v the vehicle speed, θ the steering wheel angle, a_{ACC} and a_{LCA} the status of the ACC and LCA respectively.

Secondary Task Engagement: A secondary task engagement S is defined as an interaction sequence and its corresponding glance sequence and driving sequence $S = (I, G, D)$. We consider all driving observations starting before the first interaction until after the last interaction such that $t(i_1) - t_b < t(d_n) < t(i_N) + t_b$. Where t_b represents a buffer duration. Regarding the glance sequence G, we consider all glances whose start time or end time falls in between the first and last interaction of I such that $t(i_1) < t^s(g_n) < t(i_N) \vee t(i_1) < t^e(g_n) < t(i_N)$.

Chapter 3

Data Collection and Processing

To enable the data-driven evaluation of IVISs and to generate insights into driver behavior that can support decision-making throughout the UX design process, we need to log, process, and analyze the respective data. Here, data engineering and data processing play a vital role in data-driven decision-making, even though often not visible [67]. The quality of data-driven decisions heavily depends on the quality of the data itself [178], or as Brynjolfsson and McElheran [179] put it: "Better data creates opportunities to make better decisions."

In this thesis, we analyze data from more than 100 Mercedes-Benz test vehicles. These vehicles are used for a variety of test procedures as well as for employee commuting and recreational driving. Although the vehicles are part of Mercedes-Benz's internal test fleet, no additional instrumentation in the form of sensors or telematics devices has been installed. Data was collected over the air using the *Telematics Data Logging Framework*. Although this telematics framework allows data to be collected from any modern vehicle in the fleet of Mercedes-Benz, and therefore from any customer who agreed to the data being collected, we only collected data from internal test vehicles. Due to this restriction, we consider the data used in this thesis to be naturalistic data and not natural data (see Section 2.2.2). All test vehicles equipped with the latest software architecture, a stereo camera for glance detection, and ACC and LCA technology contributed to the data collection. In the following, we present the data acquisition framework and the individual signals and their processing. This chapter is partly based on previous publications [3, 4, 6].

3.1 Telematics Data Logging Framework

Data collection and processing is based on a feature-usage logging mechanism for the telematics and infotainment system that enables Over-The-Air (OTA) data transfer to the *Big Data Platform*, where the data is processed, stored and made available for off-

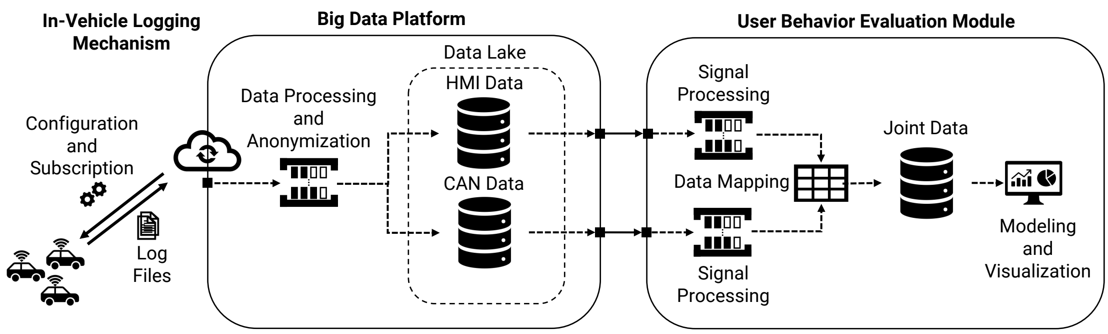


Figure 3.1: Overview of the Telematics Data Logging Framework.

board data analysis to gain insights into driver behavior and driver interactions with the IVIS. The system architecture consists of three main parts: (1) the *In-Vehicle Logging Mechanism*, (2) the *Big Data Platform*, and (3) the *User Behavior Evaluation Module*. An overview of the system is given in Figure 3.1.

The In-Vehicle Logging Mechanism is a network-enabled Head Unit (HU) application that collects data from the HMI and the Controller Area Network (CAN) bus. The HMI data (e.g., touchscreen interactions) is collected via an interface that enables communication between different applications within the HU. The CAN data (e.g., vehicle speed, driving automation status, glance data) is collected through an abstraction layer between the In-Vehicle Logging Mechanism and the CAN bus. Data collection is trip-based, meaning that when the ignition is turned on, the In-Vehicle Logging Mechanism generates a *sessionID* for that specific trip and communicates this sessionID to the Big Data Platform. This sessionID is the only identifier that links the data points of a trip.

Additionally, after turning on the ignition, each vehicle sends a request to the Big Data Platform asking if a new configuration file is available. The configuration file is transmitted over the air and specifies signals to be logged, their frequency, and their aggregation type. Once configured, data packets containing log files are sent to the Big Data Platform at regular intervals until the ignition is turned off. The Big Data Platform receives, processes, anonymizes, and stores the data in a data lake. To comply with the GDPR and protect the privacy of drivers, the data undergoes a strict anonymization process. This anonymization process truncates the vehicle identifier and modifies the timestamps of all data points within a trip by the same random value. As a result, it is not possible to link data points across trips or to link them to a specific vehicle. After anonymization and processing, each data point consists of the signal name, the payload, the sessionID, the truncated vehicle identifier, the software version of the HU, and the timestamp t.

The User Behavior Evaluation Module, developed in conjunction with this thesis, implements an Extract, Transform, Load (ETL) pipeline. This pipeline extracts the HMI and CAN data stored in the datalake, transforms it, and loads it into a datalake that makes the processed data accessible for modeling and visualization. Data transformation (i.e., data processing and sequence extraction) is discussed in detail below.

3.2 Data Processing and Sequence Extraction

The User Behavior Evaluation Module processes touchscreen interactions, driving data, and glance data. To ensure data quality and prepare the data for subsequent analysis, modeling, and visualization the data is extracted and processed at signal level. Subsequently, secondary task engagements that match the definitions given in Section 2.4 are extracted.

3.2.1 Interaction Data

For each touchscreen interaction, a JavaScript Object Notation (JSON) object is logged that contains information about the interactive UI element that was triggered, the press and release coordinates c of the fingers, and the client ID of the touchscreen on which the interaction was recorded. Since we are only interested in the driver's interactions with the center stack touchscreen, we discard all interactions on the passenger touchscreens (see Figure 1.1).

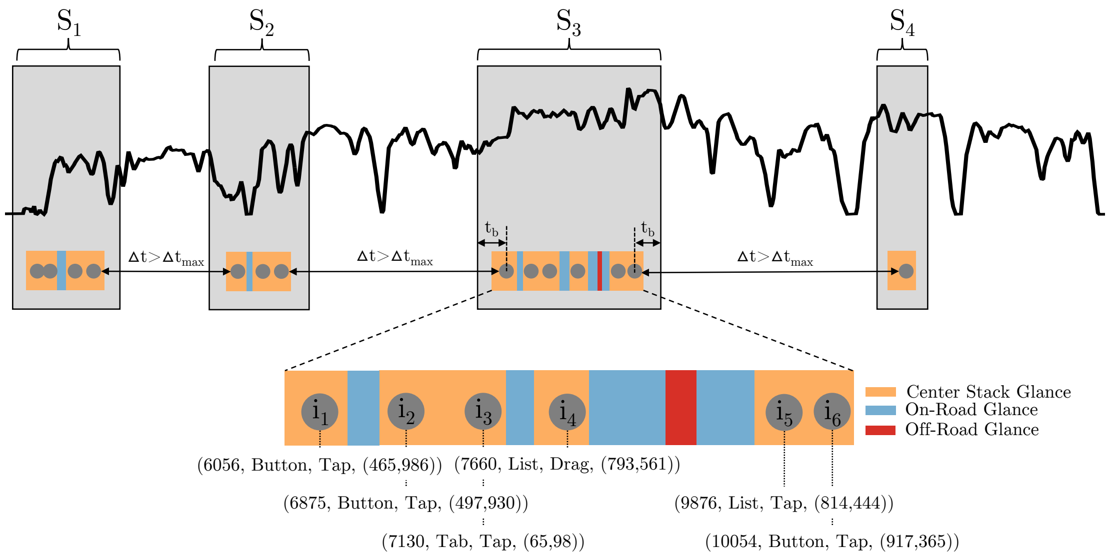


Figure 3.2: Schematic overview of how secondary task engagements S_n are extracted from driving sequences (solid black line), glance sequences (colored rectangles), and interaction sequences (gray dots).

In controlled experiments, participants are instructed to perform pre-defined tasks specified by the experimenter. In such settings, it is straightforward to map user interactions and tasks. However, in this observational setting, we do not know the drivers' intentions. Thus, we cannot infer which interactions belong together to form a goal-oriented task. One way to extract interaction sequences would be to consider all interactions that occurred during a trip. However, this would result in very long interaction sequences with dense clusters of interactions sparsely distributed over a long period of time.

Therefore, as defined in Section 2.4, we extract interaction sequences based on the assumption that drivers are disengaged from the secondary task if they do not interact with the touchscreen for more than $\Delta t_{max} = 10\,\mathrm{s}$ (see Figure 3.2). So we split the initial interaction sequence whenever $t(i_{n+1}) - t(i_n) > 10\,\mathrm{s}$. The first interaction after the split is then considered the start of a new interaction sequence.

We argue that the 10-second assumption is valid because both the distribution of interaction sequence durations and the distribution of total glance times toward the center stack touchscreen associated with the interaction sequence (as reported in Chapter 8 and Chapter 9) match well with values reported in the literature [15, 180]. For all touchscreen interactions, we infer the gesture type p (*Tap*, *Drag*, and *Multitouch*) from the press and release coordinates of the detected touch points. We also compute the distance between two touch interactions using the finger position information. Finally, each touch interaction is assigned to one of the broader element types e shown in Table 3.1.

Table 3.1: Overview of the different UI element types and touch gestures inferred from the touchscreen signal.

Category	Description
UI Element Type *e*	
Button	General buttons like push buttons or radio buttons
List	List containers used (e.g., to show destination suggestions)
Homebar	Static element containing home button, music and climate controls
AppIcon	Application icons on the home screen, used to start an application
Tab	Tab bar used to navigate between different views or subtasks
Map	Map viewer that displays a map and allows for interactions with it
Keyboard	Virtual keyboard or number pad to enter text
CoverFlow	Animated widget that, for example, visualizes album covers
Slider	Vertical or horizontal sliders (e.g, to adjust the volume)
RemoteUI	Apple Car Play or Android Auto
ControlBar	Menu controls to show context menus or popups
ClickGuard	Non-interactive background elements
Browser	Web browser
PopUp	Pop-up elements (e.g., to confirm an action)
Other	UI elements that do not fit any of the above categories
Unknown	UI elements for which the identifier is not specified
Gesture Type *p*	
Tap	A one finger touch on the screen without significant movement
Drag	A one finger dragging motion
Multitouch	A multi finger gesture

Table 3.2: Overview of the driving-related signals.

Driving Parameter	Description
v	Vehicle speed in km/h
θ	Steering wheel angle in $^\circ$
a_{ACC}	Status of the adaptive cruise control $a_{\text{acc}} \in \{0, 1\}$
a_{LCA}	Status of the lane centering assist $a_{\text{lca}} \in \{0, 1\}$
b	Status of the front passenger buckle switch $b \in \{0, 1\}$

3.2.2 Driving Data

The collected driving data consists of vehicle speed v, steering wheel angle θ, and two signals a_{acc} and a_{lca} indicating the status of ACC and LCA respectively. We also consider the passenger buckle switch signal b as driving data. Steering wheel angle and vehicle speed are recorded at a frequency of $4\,Hz$, and the remaining signals are recorded as they change.

The ACC and LCA signals indicate whether the respective system was active (1) or inactive (0). ACC automates longitudinal control and LCA assists lateral control by actively steering to keep the car in the center of the lane. Both systems work at speeds between 0 km/h and 210 km/h. An additional feature is the so-called Active Traffic Jam Assist. If both ACC and LCA are active and the driver is stuck in a traffic jam on a multi-lane road with separate lanes, the system can fully control steering and acceleration up to 60 km/h. However, according to SAE J3016 [30], it is still a Level 2 driving automation

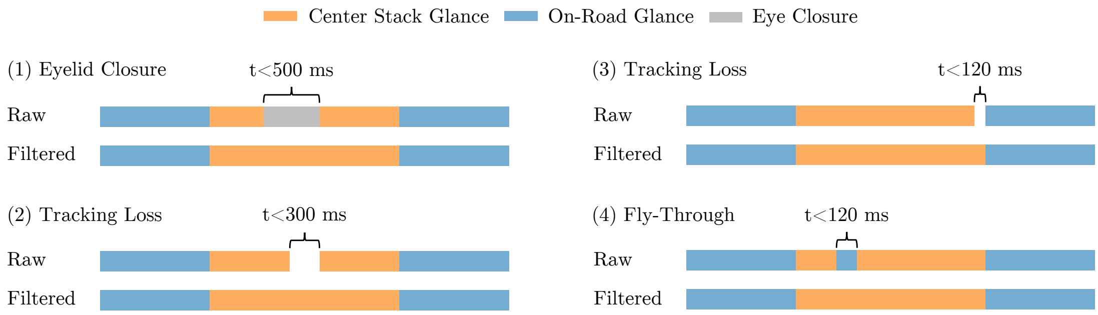


Figure 3.3: Illustration of the glance processing procedure: (1) Eyelid closures shorter than 500 ms, preceding and subsequent AOI are similar, (2) Loss of tracking shorter than 300 ms, preceding and subsequent AOI are similar, (3) Loss of tracking shorter than 120 ms, preceding and subsequent AOI are different, (4) Fly-through shorter than 120 ms, preceding and subsequent AOI are similar.

system, which means that the driver is still required to monitor the driving environment at all times. The front passenger buckle switch signal indicates whether the seat belt is buckled ($b = 0$) or not ($b = 1$).

To generate driving sequences according to Section 2.4, we extract all data relevant to a given interaction sequence. For each interaction sequence, we consider the vehicle speed and steering wheel angle data from two seconds before the first interaction until $t_b = 2\,\mathrm{s}$ after the last interaction (see Figure 3.2). This allows us to compute more stable aggregate statistics for very short sequences. Finally, we discard all sequences for which we find deviations in the logging frequency.

3.2.3 Glance Data

The glance data is collected using a stereo camera located in the instrument cluster behind the steering wheel. The eye-tracking is primarily based on the pupil-corneal reflection technique [181], which is used in the majority of remote eye-tracking devices [182]. The driver's field of view is divided into 21 different AOIs. The system continuously tracks the driver's gaze by mapping it to one of the AOIs. The true positive rate of the AOIs describing the center stack touchscreen is over 90 percent. No raw video data is transmitted at any time, as only the AOIs, eye closures, and tracking losses are logged.

We apply several processing steps to improve the data quality of the glance data. The processing is partially adapted from related research by Morando et al. [23] and ISO 15007-1:2020 [39]. We aggregate the glance information into broader AOIs (*On-road*, *Off-road*, *Center Stack*). According to ISO 15007-1:2020 [39], we consider all glances that are not directed directly at the road (e.g., glances in the rearview mirror) to be off-road glances. Since we are explicitly interested in glances toward the center stack touchscreen, we distinguish these glances from other off-road glances. Second, as described in Section 2.2.2, we extract all glances between the first interaction i_1 and the last interaction i_N of each sequence and apply various filtering steps as shown in Figure 3.3. (1) First, we filter all eyelid closures shorter than 500 ms to remove normal blinks and eyelid closures not associated with microsleeps. (2) To handle short periods of tracking loss, we interpolate gaps shorter than 300 ms if the preceding AOI is equal to the succeeding one, and (3) gaps shorter than 120 ms if the preceding and succeeding AOIs are different. 120 ms is the shortest fixation that humans can control [39] and shorter fixations are physiologically impossible. Accordingly, to remove fly-throughs, (4) we also interpolate all glances shorter than 120 ms. When glances are interpolated, the duration of the filtered glance or tracking loss is added to the duration of the previous AOI if the preceding and subsequent AOI are different (see (3) in Figure 3.3). If preceding and subsequent AOI are similar, the surrounding glances are merged as displayed in (1) in Figure 3.3.

Part I

Understanding Data-Driven Methods to Improve Automotive UX Design

As outlined in Section 1.2, the growing number of features, the introduction of large center stack touchscreens, and the impact of driving automation on driver behavior increase the complexity of designing and evaluating IVISs. The effort associated with traditional UX evaluation methods, which are mostly based on qualitative and small-scale user studies, is beyond the capacity of automotive UX design teams. As a result, they tend to neglect the principles of user-centered design. We argue that data-driven methods can make IVIS design and evaluation more efficient and can be a useful complement to traditional methods. However, we first need to understand how data-driven methods can potentially facilitate the design and evaluation of IVISs. How can large amounts of customer usage data improve UX activities? What tools, insights, or methods do practitioners need? What prevents automotive OEMs from using data-driven methods to evaluate IVISs? In Part I we present answers to these questions. We report on two studies that provide a comprehensive understanding of the role, the potentials, and the limitations of applying data-driven methods in the automotive UX design process. In addition, our research highlights which methods, tools, and insights are most needed by automotive UX experts. The findings from this chapter serve as a guide for the remainder of the thesis.

Chapter 4

The Role and Potentials of Data-Driven Methods in the Automotive UX Design Process

Context Despite the general growing awareness of the potential of big data analysis, there is a lack of research on how data-driven methods based on customer usage data[1] can support automotive UX experts in their design and decision-making process. To develop these methods, we need to understand the role that data-driven methods currently play, the challenges UXexpert face in the design process, the potentials that practitioners see in analyzing large amounts of data, and the concerns that they share. In this chapter, we present the results of a qualitative study investigating the current role of customer usage data in the automotive industry and highlight the differences with digital products. We conducted semi-structured interviews with 14 UX experts, 8 of whom currently work in the automotive industry and 6 of whom work in other industries. The interviews focused on the current state, challenges and potential of customer usage data in the respective UX design process.

Contribution Our key findings indicate that implicit feedback from customer usage data is not currently evident in the automotive UX design process. Most design decisions are made based on personal preferences and intuition of stakeholders. Our interview results show that customer usage data has the potential to reduce the influence of guesswork and biased judgments in the UX design process and can help UX experts in evaluating IVISs to make more evidence-based and user-centered design decisions. We extract the explicit needs of UX experts, present the challenges they currently face in evaluating IVISs, the concerns they share about data-driven methods, and the potential they see.

Related Publications This chapter is adapted with minor changes from Ebel et al. [1]

[1] In the original publication [1] we refer to customer usage data as field user interaction data.

4.1 Study Design

Despite the claimed potentials of using customer usage data to improve the UX development lifecycle and its success in other fields, there are indications that these potentials are not (yet) leveraged in the development of automotive IVISs. We are interested in *why* this is the case. We want to answer the following research questions:

RQ1: What is the current role of customer usage data in the automotive UX development lifecycle?

RQ2: What are the needs, challenges, and concerns in the context of data-driven UX Development?

RQ3: How can the automotive UX development lifecycle benefit from field-data-driven approaches?

RQ4: What is specific to the automotive UX development lifecycle and what can be generalized from digital companies?

4.1.1 Research Method and Interview Design

To answer the research questions we followed a qualitative approach and conducted semi-structured interviews. Before conducting the interviews, we asked the participants to answer a questionnaire regarding their demographics, background, and experience. Although we prepared a list of questions[2], we varied the order of questions to unfold the interview conversationally. This exploratory approach allows open-ended questions and engages the participants to independently address the objectives they consider important. The interview itself was divided into three parts addressing the three usability engineering lifecycle phases introduced by Nielsen [57]: pre-design, design, and post-design. Regarding each phase, we asked the participants about the methods they currently apply, the challenges they face, and the potentials they see in data-driven approaches. Each interview lasted approximately one hour and was conducted by the first two authors with always one interviewee present. Of the 14 interviews, 5 were carried out in person, one via video call and 8 via phone.

4.1.2 Study Subjects

In total, we interviewed 14 UX experts from 11 different companies, 8 working in the automotive industry, and 6 working for digital companies. We define a digital company as a company whose main product is a digital product or which has a digital product in its core business. The domains of these digital companies range from digital music services through e-commerce to telecommunications. However, we carefully selected candidates that are solely responsible for a digital product within their company. To get this broad range of perspectives inside each of the groups we applied purposive sampling [183]. We approached companies of different sizes and domains and selected candidates of various backgrounds. The interviews were conducted between October 2019 and March 2020. Since all participants are kept anonymous, they are referred to with IDs P1-P14. All

[2]The interview guideline is given here: `http://kups.ub.uni-koeln.de/id/eprint/65348`

Table 4.1: Demographic information of participants.

#	Age	Education	Job Title	Industry	# Employees	Experience
P1	20-29	Master	User Researcher	Automotive (RH)	201-500	1-4 years
P2	40-49	PhD	Technical Specialist	Automotive (OEM)	10,001-100,000	10-19 years
P3	20-29	Bachelor	UX/UI Designer	Automotive (RH)	51-201	1-4 years
P4	30-39	Master	UX/UI Designer	Automotive (RH)	201-500	1-4 years
P5	30-39	Bachelor	UX/UI Designer	Automotive (RH)	201-500	10-19 years
P6	30-39	Master	UX Marketing Specialist	Automotive (OEM)	10,001-100,000	1-4 years
P7	40-49	Bachelor	Interaction Designer	Automotive (RH)	501-1,000	10-19 years
P8	40-49	Diploma	Project Manager UX	Automotive (OEM)	10,001-100,000	1-4 years
P9	20-29	Master	Interaction Designer	Internet of Things	501-1,000	5-9 years
P10	30-39	PhD	UX Manager	E-Commerce	10,001-100,000	5-9 years
P11	30-39	PhD	Ergonomist	Telecommunications	100,000+	1-4 years
P12	40-49	Diploma	Head of UX	IT Service Provider	1,001-10,000	10-19 years
P13	40-49	Bachelor	UX/UI Designer	Apps and Software	51-200	1-4 years
P14	30-39	Master	Design Manager	Digital Music Service	1001-10,000	10-19 years

participants are currently employed in industry, with only 3 never having worked in a research context. Table 4.1 shows an overview of the demographics of the participants. In the automotive industry, it is very common that OEMs have multiple smaller research facilities or Research Hubs (RHs), where specialists work on a specific topic, decoupled from the main company. The participants did not receive any compensation.

4.1.3 Data Analysis

The first author transcribed and anonymized the audio recordings of the interviews. Afterward, the first and second authors applied a mixture of a priori and emergent coding [184] in a collaborative manner using ATLAS.ti[3].

For initial coding, both authors agreed on a set of codes based on the research questions. However, the authors were free to introduce new codes whenever they considered it to be necessary. For the coding, no special restrictions applied and each interview transcript was coded independently by the first two authors. To ensure the reliability of coding, the inter-coder agreement, according to Krippendorff [185], was calculated before the results of each interview were discussed and merged. The inter-coder agreement over all interviews is $\alpha = 0.822\,(\sigma = 0.119)$ representing a satisfactory result [185]. Newly introduced codes were reviewed by both authors and after mutual agreement, were added to the set of codes. This procedure was repeated for each interview and already coded transcripts were updated collectively by the authors. The changes introduced to the set of codes decreased after 6 interviews and no new codes emerged after 11 interviews. Therefore, we conclude further interviews will provide only a few (if any) new insights and we reached a point of *theoretical saturation* [186].

The quotes from non-English speaking interviewees were translated into English and edited for readability. Colloquial expressions were not changed to reflect the informal setting of the interview.

4.1.4 Threats to Validity

The five threats to validity in qualitative research identified by Maxwell [187] also apply to our study design. These threats describe the flaws that can occur while obtaining and interpreting the study observations. Further, the collected data might be manipulated to fit a specific theory, may it be deliberately or accidentally. Maxwell [187] argues that the researchers must preclude those threats by developing a study design that provides evidence that no “alternative hypotheses” can be made [188].

Descriptive validity refers to the threat of incomplete and inaccurate recordings. To preclude this threat all interviews were recorded and transcribed. The transcripts are annotated with timestamps such that the original conversation can be traced back during analysis.

The threat of *interpretation validity* addresses the challenge to capture the observations as intended by the participants. To avoid this threat, we used open-ended and non-directional questions. Additionally, all interviews were independently coded by two authors, and potentially ambiguous statements were discussed to identify the interpretation intended by the participant.

[3] https://atlasti.com/

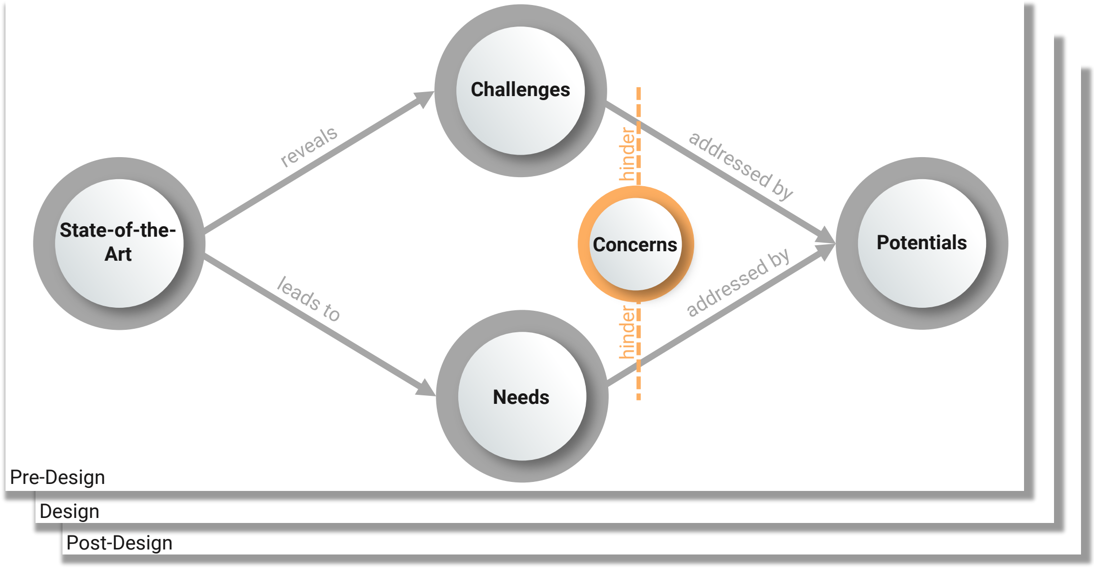


Figure 4.1: Illustration of the thematic coding model applied to the interview transcripts.

Theory validity and *researcher bias* refer to the threat that the researchers force the data to fit a certain theory they want to support or that they possess a deliberate bias regarding the participants or a certain outcome. Mitigating this threat is the fact that the study is constructed to be very exploratory, having the intention to reflect the current state-of-the-art in the industry and identify potentials. Additionally, we lowered the researcher bias by applying the introduced coding and reviewing concepts.

Reactivity describes the threat that the presence of the interviewers may influence the interviewees. This threat can hardly be mitigated but still, the authors payed attention to not influence the interviewees when conducting the interviews.

An additional threat is posed by the selection of the interviewees. We only interviewed employees of automotive OEMs, which might introduce some bias by excluding suppliers. The OEM research hubs usually act as company-internal suppliers, being solely responsible for whole systems within the car, which might add some similar perspectives.

4.2 Results

We structured the identified codes into categories and illustrate their relations in Figure 4.1.

The model shows that the reported **State-of-the-Art** reveals **Challenges** and leads to **Needs** of practitioners. Some of these challenges and needs can be addressed by analyzing customer usage data (**Potentials**). These potentials are expressed explicitly and implicitly by the participants. **Concerns** were mentioned as hindering factors. The model applies to the pre-design, design, and post-design phase.

Figure 4.2 shows the distribution of codes within each category as bars, and the number of interviews the code occurred in as numbers on top. On average we introduced 80 ($\sigma = 26$) codes per interview.

State-of-the-Art: Statements of phenomena in current practice, which reveals a challenge or leads to a need.

Challenges: Statements of problems that arise from current practices.

Needs: Statements of demands towards improving the UX development lifecycle.

Concerns: Statements of doubts that a challenge can be overcome or a need can be fulfilled.

Potentials: Statements of areas where data-driven approaches may address a challenge or fulfill a need.

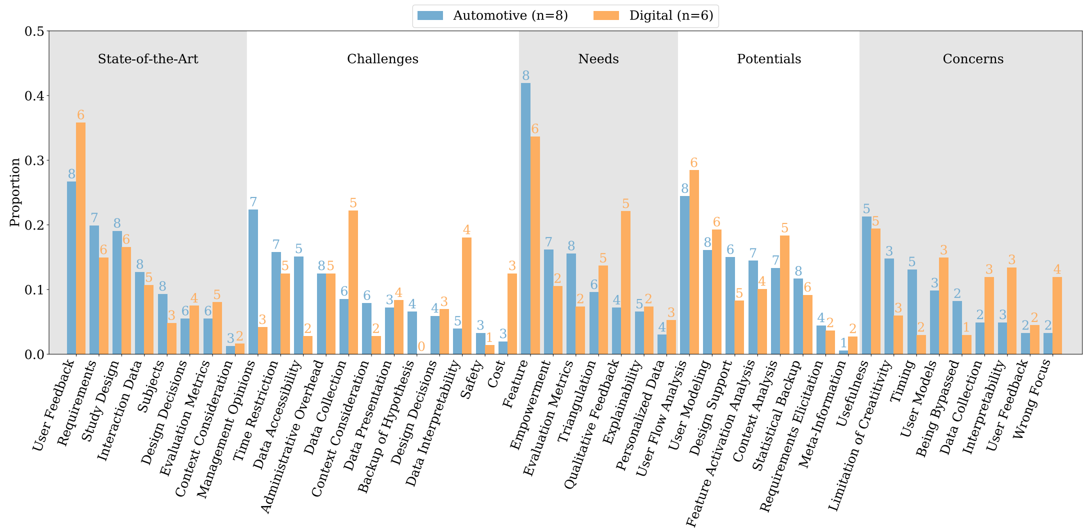


Figure 4.2: The code distribution shows the distribution of codes within each category. The number on top of each bar represents the number of interviews, per domain, in which the code occurred.

4.2.1 State-of-the-Art

In the pre-design phase, the most frequent code in the state-of-the-art category is **Requirements**. Six of the eight interviewees from the automotive domain agree that the requirement and feature elicitation is not user-oriented. P1 states: *"[. . .] at this point [the pre-design phase] we have no clue if the customer [user] is interested in this feature or not"*. Only interviewee P2 from the automotive domain confirmed that they already use some form of customer usage data in the pre-design stage by aggregating data from company cars in all markets in real-time. This data is then used, for example, to derive statements about the usage frequency of certain features to *"prioritize what [the company] actually should spend money on and what [. . .] the most important features based on usage [are]"* (P2).

Another quarter of citations in the pre-design phase are tagged as **User Feedback**, being the most mentioned state-of-the-art code over all phases (see Figure 4.2). All participants from the automotive domain but P5 mention they receive some form of user feedback in the pre-design phase. This feedback, however, is usually in the form of general market research and not really focused on the users' explicit needs or behavior. In contrast, P2 describes their rather elaborate process of analyzing user needs: *"We do that [long-term ethnographic research] by observing, interviewing, participating with people in their life, being in their homes, trying to figure out what life people are living, what are their pleasure points and pain points"*. In contrast, within digital companies, the elicitation of features and requirements seems to be more focused on the user. All digital domain participants report that the ideas in early development phases are created together with representative customers, are based on insights drawn from customer usage data, or both.

In the design phase, two-thirds of state-of-the-art citations are coded with **Design Decisions**, **Study Design**, or **User Feedback**. Regarding the automotive state-of-the-art, 6 out of 8 practitioners confirm that they evaluate their designs mainly in-house in an informal, qualitative way with coworkers and other UX experts. In contrast, P2 and P6 from the automotive domain confirm that they recruit external people on a regular basis for early feedback on their designs and ideas. Interviewee P7 describes the current state-of-the-art: *"Testing within [the company] is sort of ok if you just need to do something quickly, but if we want to verify things, it's better to get people that are not familiar with what we do"*. While all participants implement clickable prototypes for their products, these are only evaluated qualitatively. The automotive participants agree that their current process of gathering qualitative feedback on prototypes is quite advanced. At the same time, none of the automotive interviewees have direct access to or actively aggregate customer usage data. These circumstances also show automotive UX experts mainly rely on explicit feedback from users and do not leverage implicit feedback through customer usage data. As mentioned before, with the tracking of usage data in company vehicles, interviewee P2 reports the most advanced data-driven processes of the automotive participants.

In the post-design phase, 7 out of 8 automotive participants confirmed they do not get consistent and detailed feedback based on field user interaction on their product. Five of these practitioners, however, confirm that they do get feedback through market analyses and general customer surveys. Interviewee P4 describes the feedback process in the post-design stage as follows: *"At the moment, we only receive feedback through studies [the company] executes, that take weeks and months. They take the [product], test it in multiple markets with many people, and curate a [report] with the results."* In the digital domain, 4 out of 6 practitioners have implemented a process to receive user feedback for their products based on customer usage data.

Key Findings
State-of-the-Art in the automotive industry (RQ1):

- Requirements and feature elicitation is not user-centered and only supported by general market research.
- Focus is on explicit feedback; implicit feedback through customer usage data is not evident.
- Prototypes are mainly evaluated qualitatively by co-workers and in-house UX experts.

Differences to digital companies (RQ4):

- In digital companies, customer usage data is considered for decisions throughout all design phases.
- For feature elicitation and prioritization, digital companies use a mixture of explicit feedback from representative users and implicit insights from customer usage data.
- Within digital companies, insights from customer usage data are very broad and range from feature usages to sophisticated hypotheses testing.

4.2.2 Challenges

Figure 4.2 shows that **Management Opinions** are often considered a challenge in the automotive domain. In particular, in the pre-design phase, six out of eight participants from the automotive domain report that their findings from user research are not considered in the decision-making process. They argue that their proposals are often overruled by higher management even though they provide evidence through their research. Practitioners from digital companies do not experience this challenge as often. This correlates with the challenge to back up the designer's hypotheses toward user interaction with the product (**Backup of Hypothesis**). This challenge was expressed only by automotive participants. P7 states: *"There are a lot of assumptions that people make about who is driving our cars, but none of them is actually backed up with any kind of information"*. **Data Accessibility** and **Data Collection** are also mentioned frequently in the pre-design phase. Data accessibility refers to a generally insufficient availability, i.e. accessibility of user-related data within the company. P8 mentions that *"[t]here is a very strong silo mentality in companies in the acquisition of information, but also in the distribution of information"*. The fact that all citations tagged with **Data Accessibility** come from automotive participants highlights the significant deficits regarding data transparency. Considering the data collection challenges, all participants mainly refer to the challenge of collecting data as detailed as possible without violating legal restrictions. However, there are further technical peculiarities that complicate extensive data collection from users in the automotive domain. P7 states that for *"the older systems none of this existed, so we have no way of understanding what people did with it"*. Additionally challenging is the need for long-lasting architectures and the heterogeneous data processed by multiple different Electronic Control Units (ECUs). P2 exemplifies that the current architecture of their vehicle platform is not yet prepared for the kind of interaction logging needed today.

Time Restriction and **Design Decisions** are the most often mentioned challenges by all participants regarding the design phase. Six participants describe that they often lack time to dive deep into user studies or usage data. P1 explains: *"The first priority is speed. We can't work on data for two or three days"*. Considering design decisions, 5 out of 8 automotive participants see a significant challenge in evaluating their designs and prototypes with regard to the context, i.e. the driving situation (**Context Consideration**). The participants further describe that the driving task itself and the driving context affect how the driver interacts with the system. The difficulty of recreating this driving situation in a lab experiment is explained by P1: *"The difference lies in the dual-task paradigm. When you are in the lab, the interaction with the HMI is the primary task, when you are driving it is only the secondary task"*. P5 adds that dynamic driving simulators offer the possibility to model the driving situation to a certain degree but that due to high cost and low availability they are only used for very few studies. The participants from digital companies focus on **Data Interpretability** and what methods need to be applied to make reliable statements.

In the post-design phase, the challenge of **Data Accessibility** reoccurs. Three automotive participants argue that the biggest challenge after a product's release is to get customer usage data to evaluate how the product is accepted by the users. P2 states that *"one of the main challenges is to make the right data available at the right time"*. In addition to the data being available, the challenges of intuitive **Data Presentation** are discussed by the participants as well. Six participants express that, due to the amount of data, customer usage data needs to be visualized in an intuitively understandable way. P4 underlines this challenge by saying that *"an 80-page pdf with results [...] doesn't help that much because nobody wants to read through it and it doesn't motivate designers to change anything"*.

Key Findings
Challenges in the Automotive Industry (RQ2):

- Customer usage data is often not available or accessible throughout the design process due to organizational, legal, or technical restrictions.
- User research is not valued; Evidence-based design decisions are overruled by management.
- The complexity of the driving context further affects the already challenging task to create insights from large amounts of customer usage data.

Differences to Digital Companies (RQ4):

- The disparity between user insights and management opinions is less challenging for digital companies.
- Digital companies face more mature challenges in terms of integrating data in their design process rather than technical or organizational challenges.

4.2.3 Needs

The distribution of codes addressing the needs of the UX experts does not show major differences between the automotive and the digital domain (see Figure 4.2). Most mentioned for both groups are explicit demands for data-driven **Feature**s. The features are manifold and range from dashboards visualizing feature-specific clickstreams to the implementation of data-driven analyses in design tools.

In the pre-design phase, 5 out of 8 participants from the automotive domain mention the explicit need for data-driven solutions to support their hypotheses and proposals made in early phases of development (**Empowerment**). This is connected to the state-of-the-art and the resulting challenges, that personal opinions in higher management play an important role in feature elicitation and prioritization. However, participants from digital companies do not express this need in the pre-design phase. They rather emphasize the significance of **Qualitative Feedback** and the need for **Triangulation** of different data sources. The need for qualitative data is important in early ideation phases, especially for new products. P10 states: *"For a comprehensive redesign of a product you can't test A/B, you have to [. . .] test them qualitatively to see if it makes sense to implement the hypothesis"*.

Revisiting the challenges of time restrictions and decision-making, 5 out of 8 automotive participants express a need to automatically evaluate their designs based on data retrieved from field usage (**Feature** and **Evaluation Metrics**). P9 agrees that such a feature would facilitate their advances toward a user-centered design approach: *"[I]t can really help to defend my decisions. I guess, honestly, I'm always trying to defend it, not for myself but for the user"*. However, regarding automated analyses and models based on customer usage data, especially the digital domain participants express a need for explainability. P10: *"When you have some kind of magic box where I present a prototype and a magic score falls out, of course, people who are not so much into UX would ask: 'ok, but what does the box do? How does it get that number? Can I even trust it?' "*.

The needs expressed most often in the post-design phase address how to measure the acceptance of a developed product or feature by field users. The participants indicate a need for **Evaluation Metric**s that quantify user acceptance and how it changes over time. Among conventional metrics like the number of clicks or conversion rates, 3 out of 6 participants from digital companies say that it is necessary to correlate these ratings with other KPIs like profit or newsletter subscriptions.

Key Findings
Needs of Automotive UX Experts (RQ2):

- Statistical support based on customer usage data to leverage design hypotheses, feature elicitation, and prioritization.
- Tool support to automatically evaluate designs.
- Automated methods should offer explanations to facilitate interpretability.

Differences to Digital Companies (RQ4):

- In digital companies, many of the needs toward hypothesis support, feature elicitation, and feature acceptance assessment are already satisfied.
- In digital companies, there is a greater need to triangulate qualitative and quantitative data.

4.2.4 Potentials

In the pre-design phase, the automotive participants are particularly interested in the potential of **Feature Activation Analysis**, i.e. the evaluation of usage frequencies and duration. Especially for arguing against management opinions, 6 out of 8 automotive participants made statements that those analyses can satisfy the expressed need to empower them in their decisions. They explain that feature activation analyses accompanied by appropriate metrics can offer valuable insights into the field usage of features. Therefore, they can facilitate feature elicitation and prioritization. The participants further indicate that **User Flow Analysis** based on customer usage data can provide a deeper understanding of how the users behave in the current system. P6 states: *"[W]e are very good at building solutions but not always good at identifying the right problems"* and formulates the idea to *"take the personas themselves from the market research and enrich them with certain usage data that are important to understand the user journey"*.

To overcome challenges in the design phase, 11 out of 14 participants indicate that automated design evaluation methods based on customer usage data could offer valuable **Design Support**. This design support could be manifested in automated quantitative usability analyses or the extraction of usage patterns for different user groups from extensive field data. 13 out of 14 participants indicate that the usage of customer usage data for **User Modeling** could play an important role in their design process. P10 suggests using a *"model that represents a persona to automatically evaluate a prototype"*. Another recurring theme is the topic of context consideration. 7 out of 8 automotive participants see the potential to use customer usage data to analyze how the driving context affects user interactions with the product (**Context Analysis**). The interviewees argue that the context plays an important role in the automotive domain since the interaction with the environment is bidirectional. P1 states that it would be necessary to not only evaluate a feature based on its usage statistics but also on how its usage influence the driving behavior. The latter has a direct and potentially fatal impact on its environment. This critical correlation could be evaluated by matching user interactions with driving data like lane-keeping parameters.

In the post-design phase, the participants see the biggest potential of customer usage data in monitoring how features and products are accepted in the field. They argue that instant monitoring after release and an easy to understand data presentation would offer interesting insights into how often features are used and how the interaction changes over time. P1 elaborates on the direct connection to the subsequent pre-design phase: *"Requirement analysis would also mean looking at the data that was collected at the end of the last version again. This should ideally be a cycle and I see the methods data-driven analyses offer at every point in this development lifecycle"*.

Key Findings
Potentials in the Automotive Industry (RQ3):

- Insights from customer usage data can shift the elicitation and prioritization of features from personal best guesses to more user-centered decisions.
- Automated evaluation methods and user modeling based on customer usage data may offer valuable design support.
- Customer usage data can be triangulated with contextual data to investigate the impact of the driving situation on the interaction and vice versa, making evaluations less biased.

Differences to Digital Companies (RQ4):

- Most identified potentials apply to both, automotive and digital domains, but digital companies are more advanced in unlocking these potentials.

4.2.5 Concerns

In the pre-design phase, the participants express few concerns toward data-driven methods and the analysis of customer usage data. P1 and P7 do not see any benefit of the discussed methods when it comes to the early ideation phases of a product. P7 states that *"[t]hat's an interesting insight that maybe all the data-driven stuff has a bigger impact on everything where you try to optimize something in contrast to the work where the creative process is the main part"*.

The predominant concern in the design phase regards the **Limitation of Creativity** of the designers which might be caused by extensive use of data-driven analyses. This is strongly connected to the concerns in the pre-design stage, as participants from both groups see a risk to get stuck in small, iterative optimization processes. They anticipate that optimizing features based on historical data prevents thinking outside of the box to create something new. These concerns are related to the concerns toward **User Models**. P10 states that it is difficult to build a model without limiting creativity and describes it as an *"overfitting problem: the model has seen too much old data and is therefore not able to generalize when it is applied to something new"*. Further, 7 participants (4 automotive) are concerned about how to interpret the results produced by an automated evaluation method (**Interpretability**). They mainly argue that an explanation has to be provided to develop trust in automatic evaluation *"because usage scores alone produce very little insight"*, according to P12. P4 agrees by indicating: *"A score might be ok, but there should be suggestions or information on how the score is calculated and influenced"*.

There are very few concerns regarding the post-design phase. However, 5 out of 8 automotive participants communicate general doubts that legislation may prevent certain features and functions from being realized due to data collection restrictions. This especially holds for potentially person-related data, e.g. GPS coordinates of a vehicle.

Key Findings
Concerns of Automotive UX Experts (RQ2):

- Insurmountable organizational, legal, or technical restrictions prevent that data can be collected.
- Quantitative insights may not be useful in early ideation phases to evaluate volatile concepts.
- Data-driven methods may limit creativity and shift the focus to small incremental changes.

Differences to Digital Companies (RQ4):

- Participants from digital companies expressed more concerns.
- A lack of interpretability can lead to disuse of data-driven methods.

4.3 Discussion

In this section, we reflect on the needs, challenges, and concerns expressed by the practitioners and emphasize untapped potential in the evaluation and development of IVISs. We additionally relate our findings to prior published research and present methods that may benefit the automotive UX development lifecycle.

Leadership, culture, and the mindset within a company highly influence the usage of data-driven methods. Compared to digital-native companies, automotive OEMs find it difficult to keep up when it comes to the integration of data-driven methodologies in UX development. However, data-driven methods based on customer usage data can benefit the development and evaluation of IVISs. In the pre-design phase, we see great potential in generating a deeper understanding of the users and their behavior through analyses of customer usage data. Data-driven methods can facilitate decisions in early phases to prioritize features or products. Multiple approaches [91, 92, 189] that enhance the user understanding based on analyses of automatically collected customer usage data can be leveraged to unlock this potential for IVISs. However, as participants also mentioned, these approaches should be considered as an additional source of user feedback and not as a replacement for already existing methods.

In the design phase, automated usability tests [99, 190, 191] can play an important role in making the design process more user-centered and efficient at the same time. The fact that the context of use, i.e. the driving situation, is inherently contained in field data is another key advantage. Additionally, the possibility to explicitly map customer usage data with naturalistic driving data creates new opportunities in the design and evaluation of IVISs. One can, for example, predict driver distraction [192], secondary task engagement [193], or identify drivers based on their driving behavior [194]. This allows considering the complex interactions between driver, car, and environment without the costs and bias introduced by simulator experiments. This is in line with the findings made in earlier work on this topic [66, 195]. However, to provide the biggest possible value for practitioners, all automated methods should provide an explanatory component and be triangulated with qualitative user feedback.

In the post-design phase, there is a need to monitor the acceptance and usage of IVISs after deployment. Here, data-driven methods offer insights that can then benefit the next development cycle.

However, in line with the findings of Lamm and Wolff [196], automated and model-based approaches currently do not play an important role in the evaluation of IVISs. This originates from none of the interviewed OEMs having a system in place that is explicitly developed to record detailed user interactions. Current systems are yet built for different purposes and only modified to offer basic capabilities, while dedicated systems are only available for test fleets. Legacy car architectures and long product lifecycles aggravate the difficulty to implement such new systems. Additionally, strong restrictions regarding privacy and security are challenging for OEMs. However, a dedicated system for interaction logging that provides detailed and high-quality data is the cornerstone of the potentials introduced by field data-driven methods.

4.4 Conclusion

Our results show that data-driven methods based on customer usage data can have great value for the automotive UX design process and can play an essential role in making the design of IVISs more user-centered. All automotive domain experts in our study agree that there is a lack of implicit feedback through customer usage data in the UX design process. These findings coincide with the work of Orlovska et al. [66, 195]. Additionally, our results support the disparity indicated in Section 2.2 that in comparison to the automotive domain, digital domains are far ahead in exploring the potentials of customer usage data. We conclude that in order to design IVISs that meet user needs and are safe to use, it is necessary to move from predominantly explicit and qualitative user feedback, e.g. through customer surveys or studies, and feedback from small-scale simulator studies, to a combination of the former with implicit feedback through automatically collected customer usage data. Another important benefit of data-driven methods is the ability to statistically support designers' decisions, overcoming the current opinion-based guesswork often found in the automotive UX design process. Interestingly, in the automotive and digital domains, we identified a high potential for the automated evaluation of customer usage data and advanced user modeling based on interaction data for early prototype evaluation. These identified potentials will be the subject of future work to unlock the benefits offered by customer usage data. Finally, the results of this work facilitate research on data-driven methods in the automotive UX design process by focusing on the needs, challenges, and concerns that UX experts face today.

Chapter 5

Integrating Data-Driven Methods into the Automotive UX Design Process

Context The results of the interview study presented in Chapter 4 show that, due to a lack of customer insight, most design decisions in the IVIS design process are made based on the subjective assessment of UX experts or the intuition of stakeholders rather than on evidence obtained from users. This is in stark contrast to the principles of user-centered design (see Section 2.1.3). Automotive UX experts are aware of these shortcomings and are calling for data-driven support. However, several questions remain: What are the shortcomings in the automotive UX design process that prevent the application of data-driven methods? What are the specific needs of UX experts when it comes to using large amounts of customer data? How do we bridge the gap between data science and UX design? The IVIS design process is deeply integrated into the automotive product development process. It requires a multidimensional approach to understand the interdependencies between UX activities and potential organizational, legal, or technical constraints that prevent the application of data-driven methods. By synthesizing the results of four individual studies using a multiphase mixed methods approach, we address this problem from different perspectives.

Contribution This chapter provides guidance to researchers and practitioners on what actions need to be taken to integrate data-driven methods into the UX design process. It also provides unexplored and interdisciplinary research areas of interest to the academic community. We provide recommendations on what specifics from a UX perspective need to be considered when building an automotive data collection and analysis framework. To this end, we discuss the technical infrastructure and identified limitations, the current way of working, and how current, mostly qualitative methods can be triangulated with data-driven methods to make automotive UX design more evidence-based and user-centered. By combining the knowledge of the limitations that apply to the automotive domain, the needs of UX experts, the methods they use, and the potential for triangulation, we aim to bring data-driven methods and UX activities closer together to unleash untapped potential.

Related Publications This chapter is adapted with minor changes from Ebel et al. [2].

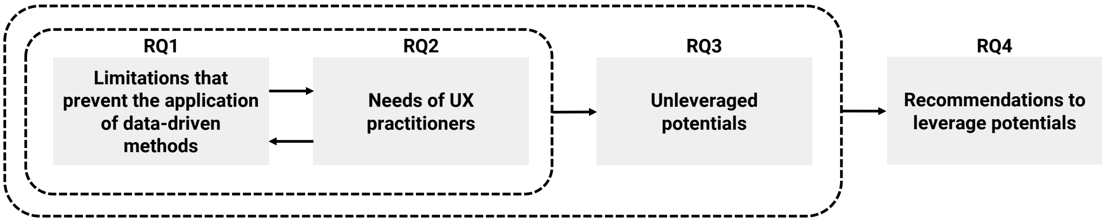


Figure 5.1: A schematic overview that shows how the individual research questions are related to each other.

5.1 Study Design

Despite the widely recognized potential of customer usage data for UCD, data-driven methods are not yet an integral part of the automotive UX design process. However, our previous results (Chapter 4) show that automotive UXexperts need data-driven support to make user-centered design decisions. Therefore, we are interested in how data-driven methods can be integrated into the automotive UX design process. Thus, we answer the following research questions:

RQ1: What are the main limitations that prevent the application of data-driven methods?

RQ2: What are the needs of UX experts with regard to the usage of data-driven methods in the automotive UX design process?

RQ3: How can implicit feedback enhance UX activities?

RQ4: What measures can improve the integration of data-driven methods into the UX design process?

Figure 5.1 shows how the individual research questions are related to each other and how we will answer them in the following sections. First, we answer RQ1 and RQ2 and how they influence each other. From the generated knowledge we then answer RQ3 before presenting specific measures related to RQ4.

5.1.1 Research Methodology

The overall design consists of four studies, two of which are interview studies with practitioners and two of which are practical investigations of vehicle data availability within OEM. Regarding the practical investigations, we applied the action research methodology [197, 198] to two studies currently being conducted in two large OEMs. The main objective of the action research methodology is to combine academic knowledge with current practical challenges [199]. While providing practical value to the client organization by introducing new methods or technologies, action research also aims to generate theoretical knowledge based on the deep and first-hand understanding the researchers gain in their interaction with the client organization [198]. The practical value of Study 1 and Study 3 lies in the data collection, data processing and data analysis methods that are introduced to the two OEMs during the course of the respective studies. The theoretical knowledge is based on the experience gained during the studies regarding the limitations, needs and potentials of data-driven methods in the UX design process. These experiences were documented during the study in the form of researcher identity memos [187]. Therefore, the action research approach builds a model of co-production between researchers and practitioners, being highly suited to evaluate the problems addressed in this work. Overall, the study explores and explains the specifics of the automotive domain in terms of using data-driven approaches to improve UX design activities. An overview of each study and its contribution to this work is provided below.

Study 1 This study consists of the design, implementation, and subsequent data analysis of a naturalistic driving study investigating the use of ADAS in different driving contexts. The study is based on data collected from 132 vehicles over a seven-month period. The purpose of this study is to identify the main limitations of current data collection processes by observing how data collection, processing, and storage are organized in practice. This study helps to identify and analyze several critical limitations regarding the use of vehicle data for user-related studies. Thus, based on the practical evaluation of two ADAS functions, this study contributes to an in-depth understanding of the underlying issues regarding vehicle data availability in one of the leading Swedish OEMs. The study design is described in detail in [200].

Study 2 The second study is an interview study conducted with the developers who designed and implemented the ADAS features that were the subject of Study 1. In this study, semi-structured interviews were conducted with the ADAS development and verification team to determine what data, and in particular what data-driven methods, are currently being used in ADAS development. All interviews were audio-recorded, transcribed, and coded separately by two independent researchers using NVivo 12 qualitative data analysis software[1]. To establish a common understanding of the coding procedure and to determine consistency and reliability among coders, both first authors of the original study [38] reviewed the codes after coding the first transcripts. After consensus was reached, all remaining interviews were coded separately by the researchers. The analysis of the interview data revealed how the data-driven evaluation process is organized and what kind of data and methods are used during the development, verification, and post-design phases. Several critical issues were identified and mapped to the different development stages. A detailed description of the study can be found in [38]. The results of this study reveal shortcomings in the effectiveness of data use, suggesting that improvements to the current data-driven approach are needed.

Study 3 In Study 3 (Chapter 4), we elaborate on the current state-of-the-art of data-driven methods and the utilization of customer usage data in the automotive UX design process. We reflect on the needs practitioners formulate toward data-driven solutions, on the concerns they share, and on the potentials they anticipate. To put the results into perspective, we conducted semi-structured interviews with UX experts from the automotive domain (N=8) and digital domains such as app or web design (N=6). The interviews were audio-recorded, transcribed, and anonymized before they were coded in a mixture of a priori and emergent coding using ATLAS.ti[2]. The identified codes were

[1] https://www.qsrinternational.com/nvivo-qualitative-data-analysis-software/about/nvivo
[2] https://atlasti.com/

structured into five categories. The relation of these categories is described in a thematic coding model. This study provides insights into the current role of implicit feedback through natural usage data, the peculiarities of the automotive domain, and the value data-driven analysis can have for automotive UX design. Additionally, the study leads to a deeper understanding of automotive-related limitations and builds the foundation for further investigation on how those limitations might be overcome. The study design and outcome are precisely described in Chapter 4.

Study 4 This study is a practical investigation conducted as part of this dissertation and relates to the development of the User Behavior Evaluation Module introduced in Chapter 3. In this study, a framework for analyzing user behavior based on natural and naturalistic event sequence data and driving data is developed and implemented. By combining driving data and user interaction collected via the Telematics Data Logging Framework introduced in Section 3.1, we evaluate the bidirectional dependencies between driving behavior and interaction with IVISs. Similar to Study 1, the action research methodology is adopted and applied. Based on the observation, evaluation and critical analysis of existing methods and current practices within Mercedes-Benz, we extracted valuable information about the current applications and limitations of data-driven approaches in the automotive UXdesign process.

Data collection in all relevant studies was conducted with the signed consent of the participants. The collection, processing, and storage of the collected data were carried out in accordance with the European GDPR, which means that the confidentiality of the data storage and the anonymity of the participants' identifiers were strictly maintained.

5.1.2 Integration and Triangulation of Study Results

In this work we adopt a multiphase mixed methods approach [69] and modified it to fit our purpose (see Figure 5.2).

Study 1 and Study 2 were both performed in cooperation with a large Swedish OEM and form the two distinct interactive phases of Study A, using an *explanatory sequential mixed methods* design [69]. The explanatory sequential mixed methods design has two distinct phases in which the action research approach precedes the qualitative interview study. In Study 1, the implementation of the design for collecting and analyzing quantitative data in a naturalistic driving study revealed several limitations and peculiarities in the company's data-related processes. These findings were then explored in more detail in a qualitative interview study with practitioners (Study 2). This study explored the limitations associated with the use of data-driven methods. The triangulation of the two studies enriches the action research findings with practitioners' insights and explanations, and helps to better understand the root causes of practical limitations.

In contrast, Study 3 and Study 4 were performed using an *Exploratory Sequential Mixed Methods* design [69]. According to this design approach, the interview study (Study 3) first explores the practitioners' needs, challenges, and concerns, which are then used to derive insights toward the practical implementation of data-driven methods for user behavior evaluation. The consecutive quantitative case study (Study 4), aims to integrate data-driven methods and tools into the UX design process of an OEM. The methods should meet the needs of the UX experts and leverage the potentials identified in the preceding interview study.

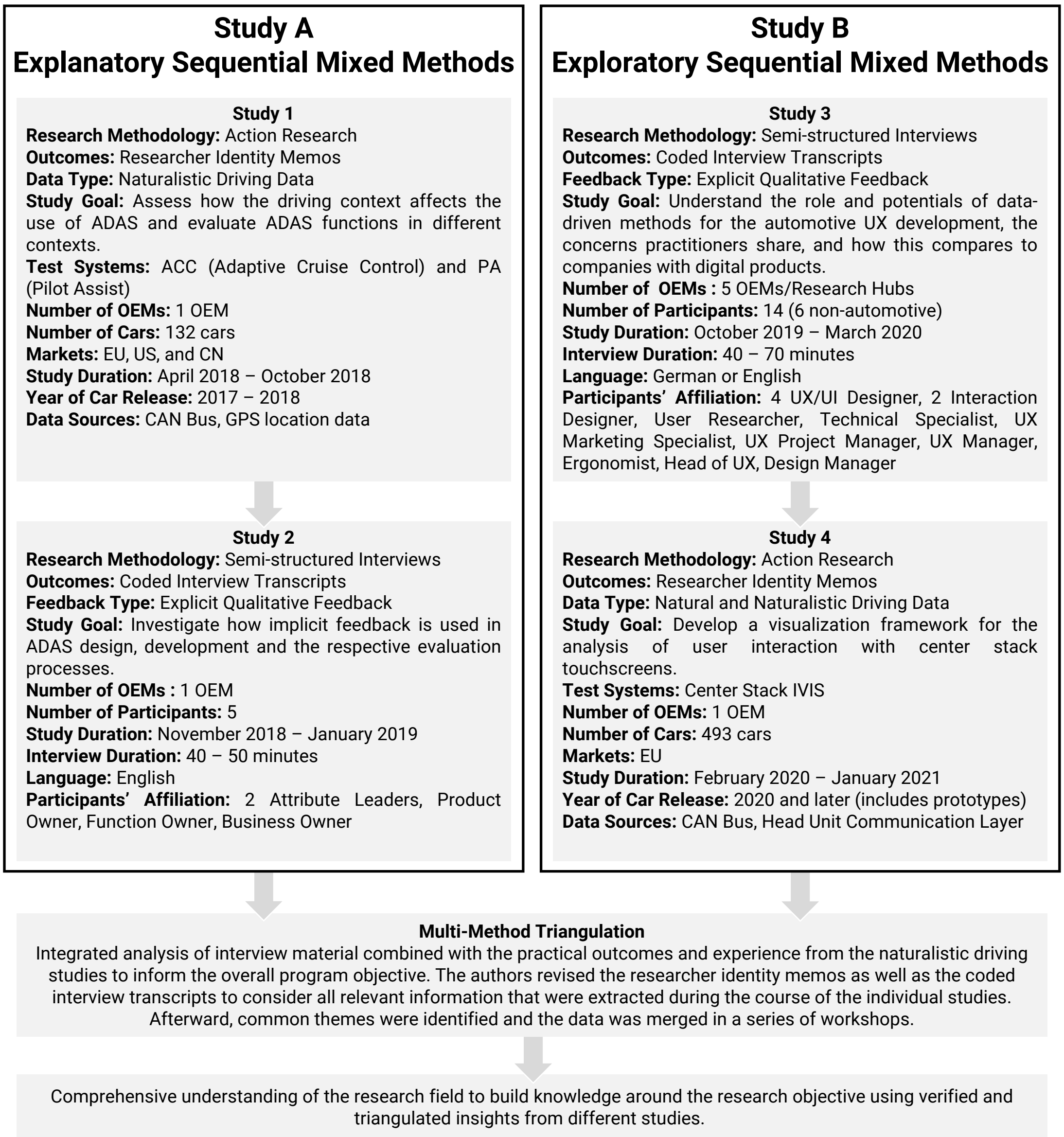


Figure 5.2: Illustration of the multiphase mixed methods approach that we used to synthesize the results of the individual studies.

Despite the parallel design of the explanatory sequential mixed methods approach (Study A) and the exploratory sequential mixed methods approach (Study B), all four studies are used to complement, enhance, and validate each other's results. For example, whereas Study 2 reveals very detailed insights, its main limitation is that it was conducted based on the practitioners' input from only one OEM, which prevents the results from being extrapolated to the whole automotive area. Thus, Study 3, which compares different automotive and non-automotive perspectives, is used to validate the results of Study 2. Simultaneously, since Study 3 does not delve as deeply into the technical details, it can be used to identify whether the limitations of Study 3 also apply to other OEMs or digital companies in general. Additionally, although Study 1 and Study 4 provide very detailed insights from working with the respective OEMs, they approach the research objective from different perspectives. Study 1 deals with the execution of a naturalistic driving study and the subsequent data analysis, and Study 4 deals with the collection, processing, and analysis of natural data.

In the first step of data triangulation, it is necessary to determine which insights can be provided by which study. While the interview studies focus more on the problems and requirements of practitioners who work directly with design artifacts (UX designers, software developers), the action research approaches shed more light on the specifics from a data science or data engineering perspective, and also bring in insights from discussions with legal and management.

In order to compare and integrate the results of all studies, the data generated were put into the same form. For the interview studies, the authors reviewed the coded raw data and extracted all limitations, needs, and opportunities mentioned by the participants. Individual statements on similar points were grouped under a common theme. The same procedure was applied to the data extracted from the researcher identity memos, which were the results of the action research methodology. A series of workshops was organized to integrate the results of the individual studies. During the first workshop, the two first authors created a mapping between the different themes to identify which points were validated or strengthened by another study. In a second workshop, the first three authors discussed the results of the first workshop and decided on the most relevant points for the UX design process. As a result of this work, a common understanding of the state of the art of data-driven methods in automotive UX was derived.

The multiphase mixed methods design expands the scope of previous research by addressing the research questions from multiple perspectives. The chosen study design allows for a comprehensive understanding of the constraints that apply to the automotive industry. It also helps to understand what data-driven methods need to be developed and how they need to be integrated to enable UX experts to leverage large amounts of customer usage data.

5.1.3 Threats to Validity

Being a joint work that combines different studies, the threats to validity of the individual studies apply to this study as well. However, a differentiation between the different types of studies has to be made. Study 2 [38] and Study 3 (Chapter 4) are qualitative user studies. Thus, Maxwell's five threats to validity [187] apply. Maxwell elaborates on the flaws that can occur during study execution and data collection, and on the threat of deliberately or accidentally manipulating the collected data to fit a certain theory. To eliminate those threats, a study must be designed such that no *"alternative hypotheses"* can be derived [188]. The individual threats and how we address them are listed below:

Descriptive validity concerns the threat of inaccurate and incomplete documentation. We have addressed this threat by recording and transcribing all interviews. Furthermore, we annotated the transcripts with timestamps such that the original conversation is easily accessible during analysis.

Interpretation validity refers to the threat of capturing the observation as intended by the interviewees. To preclude this threat, we used non-directional and open-ended questions. Additionally, the transcripts were coded independently by two authors of the respective works, and statements that could be interpreted in different ways were discussed to interpret them as intended by the interviewees

The threats of *theory validity* and *researcher bias* refer to researchers forcing the data to fit a specific theory or being biased toward the participants or a potentially desired outcome. To mitigate this threat, both studies were constructed as exploratory studies, intending to reflect the current state-of-the-art in practice. Furthermore, the coding and reviewing concepts that were applied are also intended to reduce the impact of the researcher bias.

The threat of *reactivity* occurs when the interviewees are influenced by the presence of the interviewer. Considering the chosen study setup, it is hardly possible to mitigate this threat. However, by paying attention to not influencing the participants and not leading the interviews in a certain direction, we tried to eliminate this threat as far as possible.

Another threat that applies in particular to the action research methodology, and therefore to Study 1 and Study 4, is a potential lack of objectivity and researcher bias. Petersen et al. [201] argue that the best way to reduce this bias is to involve multiple researchers and to collaborate with different practitioner groups. By involving researchers from different research institutes, all collaborating with different OEMs, we try to mitigate this threat as much as possible.

A final threat that applies to all four individual studies is the threat of *selection bias* [202]. All the information in this study is derived by working with, or talking to, UX and data science experts from a selected set of OEMs. For this reason, the statements cannot be generalized for all automotive OEMs, since the maturity in which data-driven methods are used in the UX design process varies between OEMs.

5.2 Results

In order to answer our research questions, it is necessary to take several peculiarities of the automotive domain into account, be they of a legal, technical or organizational nature. The methodology described in Section 5.1.2 allows to examine the given objective from different perspectives. Thus, we are able to make differentiated statements about the current limitations, desires, and potentials of data-driven methods in the automotive UX design process.

The results of the interview studies 2 and 3 show that the use of data-driven methods varies depending on the OEM, but also within the different phases of the respective product development process. However, since most automotive OEMs have similarities in their organizational structure and development processes, it can be assumed that the derived artifacts exist in other OEMs as well. However, the extent to which these findings can be applied may vary between OEMs.

In the following section, we discuss the limitations that prevent the application of data-driven methods and the needs of UX experts regarding the use of vehicle data. The findings are drawn from the studies presented, and the superscripts (S1,S2,S3,S4) in the section headings indicate the studies on which each statement is based. This is followed by a discussion of how the use of implicitly collected data can improve UX activities, and recommendations on how to better integrate data-driven methods into the UX design process.

5.2.1 Limitations That Prevent the Application of Data-Driven Methods

In the following, we present general limitations in the automotive software development and their consequences for the application of data-driven methods in the UX design and in the product development lifecycle (RQ1).

Automotive Software Platforms Are Not Designed to Support the Growing Needs of Data Logging (S1,S2,S3) Most automotive software platforms are not (yet) designed to meet the dynamically changing data availability requirements introduced by the rapid evolution of data-driven methodologies. Due to the high costs associated with developing a new automotive platform and software architecture, most traditional automotive OEMs choose to incrementally extend their legacy platforms. Thus, currently available data logging systems are developed as interim solutions. This results in several shortcomings concerning the formulated needs for data-driven support. According to published research, few OEMs have a logging system *specifically* designed to analyze usage data and derive detailed metrics from user interactions. However, recognizing that much progress remains unpublished, assumptions can only be made about the logging infrastructure of some OEMs.

Consequence: Most of the available data is extracted from CAN and FlexRay buses and relates to system performance data. However, the logging of user interaction data is less developed. Signals generated within specific units, such as the infotainment unit, are still limited. As a result, the data currently available to UX experts is limited in detail, quality, and consistency, and therefore poorly suited for state-of-the-art data-driven user behavior analysis.

The Product Lifecycle Is Long (S1,S2,S3) The product lifecycle in the automotive domain is long compared to digital products or other consumer products (e.g., smartphones) [203]. The ability to make changes to hardware or hardware-related signals after a vehicle is released is further limited due to the stage-gate approach. This delays the introduction of new digital technologies in vehicles that are already on the market. New technologies can, therefore, often only be introduced in the next generation of cars. An interviewee from Study 2 adds that they *"[...] specified [the data] a couple of years before the first vehicle went to production"* and further argues that it is very difficult to answer any new research question that occurs afterward, as this may require data that was not part of what was defined in the very early stages.

Consequence: The long product development lifecycle and the low flexibility contradict the fast-changing needs regarding UX design. Newly introduced data points, needed for either the development of new applications or UX analysis, are often provided with significant delay. This leads to a slower development of digital technologies compared to other digital domains.

Data Is Distributed Over Different Subsystems (S1,S2,S4) A car is a complex product, consisting of a multitude of systems, subsystems, and functions exchanging data to enable communication [204]. Often UX experts are in need of data generated by subsystems such as the infotainment system, the body and comfort systems, or the powertrain system to triangulate driver-system behavior relevant for the overall UX. One participant (Study 3) describes that the system complexity makes it hard to answer questions, that in themselves are not very complex: *"We wanted to measure how many times someone opened the window. That's a really difficult problem since you have to go through all the physical wiring and switches, so we are not certain on how to get that information"*. This example illustrates the lack of centralized databases and shared documentation that describe and organize the signals needed to design, develop, and evaluate IVISs.

Consequence: Due to the lack of a process that collects, evaluates, and orchestrates all available data points, UX experts often do not have access to potentially relevant data or its description. In addition, current databases often contain duplicates of signals resulting from the parallel development of different IVISs. These signals are often poorly described and knowledge of the interdependencies between different signals is not available. This can lead to incorrect assumptions that affect the validity of the data and the systems that use the data.

Access to Components of Suppliers Is Limited (S1,S4) The car consists of a variety of software and hardware systems that are often developed independently by external suppliers [203]. These systems are often black boxes with no access or ability to modify the code base.

Consequence: The outsourced software development introduces subsystems that can not be updated by the OEMs, making the transparent and consistent documentation of signals difficult. As a result, data scientists and UX experts struggle to derive and introduce new user-related signals from already implemented legacy systems. Inconsistent documentation and missing data can reduce the reliability and usefulness of data-driven methods.

Strict Data Protection Regulations and the Associated Internal Processes Limit Data Collection and Utilization (S1,S2,S3,S4) Advances in automotive software, intelligent applications and data-driven solutions also bring new security and privacy challenges. In particular, personal data processed in the cloud must be handled without violating the data protection regulations of the respective countries. According to a comparative analysis conducted by Voss and Houser [205], the United States and the European Union define and understand personal data differently. The *Protected Personally Identifiable Information* in the United States contains less information than the similar concept of *Personal Data* in Europe. For example, some pseudonymized information may be considered impersonal in the United States, while according to the European GDPR, the same information would be considered sensitive. China did not have a privacy law until recently. Today, China is working on building a data protection system through legal adoption and transformation of both EU and US laws [206].

Consequence: Strict regulations, particularly in Europe, restrict the collection and processing of personal data. This applies to applications that rely on the use of personal data, as well as to analyses that must be performed on personal data. In addition, it is often necessary to go through a complex legal process to make a recommendation within the OEMs as to whether certain data points are considered personal data or not. One of the UX experts interviewed in Study 3 adds that *"[...] when it comes to sensitive data, you have completely different security requirements. This means that you have to go through different audits which often critically impact the time schedule"*. While this process is indispensable and the UX experts are aware of it, they complain that it is too time-consuming, non-transparent, and also delays the processes and evaluations of non-personal data. In addition, strict data protection regulations and inadequate processes within OEMs make it difficult to obtain data from customers in the field. A participant in Study 2 states: *"we are only able to do this [i.e., data-driven evaluations], in a fairly easy way if we have access to company cars [...] because it would be very tricky to log such data from [real] users"*. As a result, qualitative data collection still serves as the main resource in user-related studies.

Hardware, Software, and UX Design Activities Are Poorly Aligned (S1,S2,S3,S4) Physical and digital parts of in-vehicle systems are often developed in parallel. Whereas the hardware of a subsystem does not change after Start of Production (SOP), software applications that build upon those subsystems are continuously developed and new UX evaluation needs constantly arise. These new applications often require new data points that were not considered at the beginning of the (hardware) system development. Another common issue is are the late specification and missing requirements from the UX departments concerning data that should serve evaluation demands.

Consequence: Poor coordination across development teams in the early stages of Product Development (PD) results in data requirements not being communicated in a timely manner. This often results in the unavailability of required data points in later stages of product development and slow development of user and contextual data.

The Possibility to Make Major OTA Updates Is Missing (S4) The car has always been a technical product, and changing requirements, physical interfaces, or functionalities after the car is released was neither necessary nor intended. In digital domains, however, practices such as A/B testing or canary releases are state of the art and considered indispensable for user-centered development [207, 208]. However, these practices require the ability to perform centralized remote updates to dynamically test new designs, fix identified bugs, and compute UX measures in real time. Verified design ideas or fixes can be deployed to production immediately. Some emerging automotive competitors have already implemented solutions [209, 210] that are available in production vehicles. However, despite the active development of such systems, most traditional OEMs are not (yet) able to make major software changes via OTA updates. In addition, the high demands toward functional safety increase the difficulties associated with online user testing. To cope with the complex interdependencies within IVISs and to ensure performance, the released vehicle is usually locked for any changes. Any further design changes are pushed to the next generation of cars. In contrast, web applications practice A/B testing of design ideas on real users, and their software allows for remote updates to fix any bugs that are discovered.

Consequence: IVIS updates remain inflexible and cannot be easily and dynamically changed based on customer feedback throughout the product lifecycle.

Looking at the above list, it is noticeable that most of the limitations are due to the specifics of current automotive product development processes. Current automotive practices, regulations that apply, priorities that are set, methods that are used, and the general vision regarding the UX design of digital products all affect how in-car solutions such as IVISs are developed today. Currently, technology-driven development is often prioritized over user-centered development because it is more obviously associated with driver safety and the reputation of the OEM. UX design comes as an important but secondary task. As a result, the data management solutions developed are more focused on the satisfaction of functional requirements rather than the data requirements introduced from the UX design side. This often leads to limitations in the design of studies based on implicitly collected data. Due to missing or poor quality data, study designs often need to be modified, resulting in study designs that do not fully fit the original research purpose. Furthermore, not all limitations are due to technical feasibility. Many OEMs still lack strategic planning for data development of user-related and contextual data. For example, user interaction data, such as clickstream data, is commonly used in the daily business of digital companies, but is still not used in many advanced automotive companies.

5.2.2 Needs of UX Experts with Regard to Vehicle Data Utilization

In this section, we present the needs of UX experts related to the use of implicitly collected data (RQ2). While some of the needs are directly related to the limitations already presented, others describe explicit needs that are independent of the current shortcomings.

Detailed Quantitative User Behavior Insights (S1,S2,S3,S4) In addition to the current mostly qualitative research, automotive UX experts need detailed user behavior data collected from multiple different sources to get a detailed picture of how people interact with IVISs. One interviewee (Study 2) emphasizes this by stating: *"[w]hat we lack knowledge about is how the real customer uses the function. That is what we must be better at"*. The data collected should be detailed enough to answer questions about specific usage patterns and usability metrics in addition to questions about the frequency and context of use. Combining different data sources is also important. For example, UX experts want to correlate interaction data with contextual data that provide insight into the driving situation.

Data Transparency (S1,S2,S3,S4) In both the interview studies and the practical investigations, one of the main needs expressed by the UX experts is the need for transparency in the data collection and processing activities. In Study 3, one participant describes a common problem as being that *"[t]here is a very strong silo mentality in companies in the acquisition of information, but also in its distribution"*. The respondent further elaborates that this leads to valuable data remaining unused. This coincides with the fact that in all studies the need for data documentation that includes all datapoints from all data sources that are available within the company is expressed. Furthermore, detailed signal documentation, technical and legal requirements giving insights about how the data is collected, processed, anonymized, and for which purposes it is supposed to be used are required. The participants argue that this is necessary to ensure that each signal is used to its full potential.

Continuous User Feedback (S1,S3,S4) To enable a more user-centric way of software development, the UX experts express the need to continuously collect, analyze, and incorporate user feedback into the design process of IVISs. They argue that the immediate and continuous feedback provided by methods such as A/B testing is needed to make data-driven and evidence-based design decisions rather than decisions based on the gut feelings of individuals or outdated market research results. One Interviewee from Study 3 states: *"I would say that the best way would be to make sure that we can do A/B comparisons directly in the cars, like other companies and online businesses do it. The customer doesn't really know that they have been updated with new functions and we can figure out which functions are best by trying different versions in different cars from different customers. So real-time evaluation with real customers of different types of function"*.

Triangulation of Qualitative and Quantitative Data (S1,S2,S3) Traditionally, UX research in the automotive domain is more qualitative than quantitative and UX researchers mostly use only qualitative approaches. However, both qualitative and quantitative data can enhance the UX activities, since both types of data contribute to a different type of knowledge [211]. One participant in Study 3, for example, expressed the need to enhance personas with quantitative evidence. This would allow them to map the personas' qualitative findings about who the target customer is with quantitative findings about how this group of customers interacts with the system. The general need for triangulation is further highlighted by other automotive UX experts stating that quantitative data might be the right choice to pinpoint a problem, but qualitative methods are still needed to further understand the problem [211]. One UX expert (Study 3) states that *"[w]ith quantitative data, we have a starting point, a trigger that tells us, let's look into this. But the quantitative data alone doesn't provide the answer to why something is happening"*. This underscores the need to combine qualitative and quantitative data to gain more detailed user insights.

Personalized or Pseudonymized Data (S1,S3,S4) Personalized or pseudonymized data is important for the development of intelligent in-vehicle applications, or for in-depth analysis of how different user groups interact with the system. For example, since a vehicle is often a shared product, personalized or pseudonymized data is needed to differentiate between different drivers that use the same car [212]. The same applies to the development of personalized services and interfaces, where the design is highly dependent on personalized driver responses to proposed solutions.

Tool and Knowledge Support (S3) Although automotive UX experts are aware that data-driven approaches can support their advances toward user-centered design, practitioners often struggle to work with quantitative data and machine learning approaches [213]. This is due to a lack of available methods, tools, and competence. Since the main task of UX experts is to deal with the design and evaluation of IVIS, there is a need for tools and methods that support them in analyzing the large amount of data that is generated by modern cars. Therefore, methods that automatically visualize data insights, calculate usability metrics, or evaluate designs based on large amounts of usage data, are needed. One interviewee explains that it would be helpful *"[...] if we could create models from user data, for example, one could directly integrate a user model into a sketch tool. Then, when creating a design it is directly evaluated against a user model"*. However, with regard to automated analyses and models, the UX experts state that such methods should also provide an explanation module so that predictions or proposed decisions can be put into perspective.

Data Visualizations (S3) Data-driven evaluation methods aim to provide UX experts with additional information so that they can make the best possible decisions to optimize the UX. To do this, UX experts express the need for intuitive data representation. They state that due to the large amount of data and the high number of different features, the data needs to be presented in an easy to understand and intuitive way. In addition, the experts argue that the information must be directly accessible without further processing. The need for fast data access is emphasized by one participant (Study 3) who describes that *"[...] it's not ideal if we always have to go to another department and say 'can you prepare this for us?' and then they say 'yes, you'll have it in a week', which of course isn't the point. It would of course be good to validate our hypotheses quickly ourselves"*. While traditional usability metrics such as average time on task or completion rates are easy to interpret, more sophisticated methods, such as those based on machine learning, should provide an explanatory component. One of the interviewees argues that while it might be interesting to get a design score for a prototype screen, the real value would come from being able to say which factors of the design specifically influenced the score. Current post-release evaluations also often take too long, which means that they are no longer of real interest once they are communicated to the UX experts.

5.2.3 How Can Implicit Feedback Enhance UX Activities?

Having introduced the limitations that prevent the application of data-driven approaches and the explicit needs that UX experts formulate towards the use of implicitly collected data, we answer the question of how implicit feedback can be used to improve qualitative UX activities (RQ3) in the following.

Data-Driven Personas The goal of the pre-design phase is to understand the target user population and derive a clear product definition. Currently, most of the UX activities in the pre-design phase are based on explicitly collected qualitative feedback. In the knowledge generation step (see Figure 2.1), the task of understanding the target group and identifying the user needs is mostly based on market research and customer surveys. One common approach to understanding who the customer is, is the persona technique. Cooper [214] defines a persona as an archetypal user, representing an underlying customer or user group. Personas are used to group similar users into a superordinate group to help decision-makers understand the customer needs [215]. Personas are typically manually created using qualitative approaches such as ethnographic field studies and interviews [216]. Therefore, manual persona generation is costly, the collected data is not directly related to the user's behavior [217], and personas tend to expire as soon as customer behavior evolves and changes [218]. Data-driven personas based on different kinds of customer data [218, 219, 220] do not only tackle the shortcomings of qualitative persona generation but aim to connect abstract personas to customer usage data. Therefore, data-driven personas can enhance the strategy and planning phase in the automotive area. Implicitly collected data retrieved during car usage can be used to generate insights on the driving preferences of different customer groups and the preference such groups have toward features such as automated driving functions, comfort, or entertainment functions. While data-driven personas might not replace the currently used personas, we argue that the triangulation of both is a promising application to create a more detailed picture of the customers. Whereas implicitly collected data retrieved from simulator studies or naturalistic driving studies can also be used to build data-driven personas, natural data has the advantage that it contains data from the whole user base and is collected continuously. It is, therefore, possible to dynamically adapt personas when changes in customer behavior take place.

Context-Dependent Evaluations Since the drivers' user experience is strongly influenced by the current driving and traffic situation [11], the designers need to understand the context of use in which the interactions occur. In the pre-design phase, no fully functional or physical prototypes exist that can be used for such evaluation purposes. However, by analyzing either naturalistic or natural data from the already existing system, it is possible to derive meta-information about the driving context, and even take into consideration the differences across markets, such as road infrastructure, traffic, and driving culture [221]. For example, aggregated data can provide insights into the length of trips, the number of trips per day, the time of day customers use their cars, or the routes they take. This information can be triangulated with the results of general market research to create a more detailed picture of how and in what context the current product is being used. In the post-design phase, implicitly collected data also has the potential to support the evaluation of driving-related functions such as automated driving. Knowing in what kind of situations driver activate or deactivate functions and how take-over requests are handled can improve post-design evaluation. Unintended or unexpected user behavior can be identified, and severity assessments of system misuse can be conducted. To date, data-driven methods for contextual monitoring have not been fully developed. Combining telematics data with external databases, traffic, weather applications, social media services, or collecting data from an in-vehicle camera, typically used in qualitative studies, is currently the most common way to assess driving context. However, the analysis of such data is time and resource consuming and raises privacy concerns. Multiple studies [222, 223, 224, 225, 226, 227] indicate great potential of implicit feedback for automated driving event recognition in real-time. The automated process of context analysis based on implicit feedback will help UX experts conduct context-aware evaluations and better understand driver choices.

Evidence-Based Feature Elicitation The feature and requirement elicitation in the automotive domain is currently mostly based on general market research and decisions are often made based on the gut feeling of decision-makers (see Chapter 4). One interviewee (Study 3) argues that *"[w]e shouldn't just carry things over for the sake of carrying things over, we should evaluate if those are actually useful things for the user. I think that's why we still have SD cards and USB Input in the car. They [decision makers] don't know if people are using it"*. UX experts often feel that their findings from qualitative or small-scale empirical user studies are overruled by management, based on the underlying assumption that they are not representative of the general user base. Insights from natural or naturalistic data can therefore be used to support and validate their hypotheses. Feature usage analysis can be used to prioritize features within the system. Analysis of clickstream or driving data can highlight current usability issues that need to be addressed. In addition, usability metrics derived from the current system can be used as input when setting usability goals for a new release. Therefore, the authors argue that triangulating qualitative research with quantitative data insights can help shift requirements and feature elicitation from personal best guesses to more objective decisions.

User Flow Visualizations The main goal of the design phase is to derive a usable implementation that can be released [57]. During the design generation and realization phase (see Figure 2.1), design ideas are gathered and initial wireframes and sketches are drawn and evaluated. While idea generation is a highly creative process, data-driven methods have the potential to help UX experts make the most appropriate design choices. The data collected in today's vehicles, which allows conclusions to be drawn about current user behavior, can serve as a source of inspiration. To realize the full potential of this data, it is important to provide designers with visualizations and tools that allow them to efficiently analyze user interaction data. Several different methods, such as Sankey diagrams [96, 97], Outflow [88] or MatrixWave [93], have proven to be efficient for many different analysis tasks and help designers to find unintended or unexpected user behavior, which in turn can be used as inspiration for new design ideas. These approaches aggregate large amounts of event sequence data and are therefore well suited for visualizing data collected in large naturalistic user studies or natural data collected from the entire user base.

Automatic Design Suggestions In addition to analyzing and visualizing user behavior data to support designers in their idea generation process, there are several approaches that automatically generate design suggestions based on different types of data. For example, Gajos et al. [113] propose *Supple*, a system that renders interfaces based on device constraints and user traces that are used to adapt the interface to specific usage patterns. Another example of a method that makes automatic design suggestions is presented by Bailly et al. [228]. Their approach makes suggestions on how to structure menus based on an adapted search-decision-pointing model used to predict selection times of menu items.

Model-Based Evaluation of Early-Stage Prototypes After design generation, wireframes are transformed into prototypes of different fidelity that need to be evaluated. In Study 3, automotive UX experts report that early design prototypes are mostly evaluated qualitatively by in-house experts or in small user studies. While evaluations with experts can provide important insights, they are not suitable for evaluating metrics such as time on task or glance behavior. However, feedback on metrics such as time on task or glance behavior are crucial for the final system and need to be evaluated as early as possible. Computational models of user behavior allow automatic evaluation of early-stage prototypes and can give valuable feedback even before a user study is conducted. Multiple approaches exist that allow predictions to be made for various metrics such as time on task [165, 167, 229, 230] or glance duration [170, 172, 173, 231]. For example, Large et al. [172] propose a method to model the visual demand of IVISs when used concurrently with driving. Their approach is based on an information-theoretic model, for which the dependent variables have been identified in a simulator study. Whereas current work is based on rather small amounts of data, generated through lab experiments, the use of big data for such prediction tasks holds great potential. Approaches based on a large amount of naturalistic or natural data allow the application of dedicated machine learning algorithms. Those applications have already proven to surpass the prediction accuracy of relatively simple regression methods in other automotive applications [116]. A general advantage of applying modeling methods to natural data is that, on the one hand, the entire user base is covered and, on the other hand, continuous data collection also implies continuous improvement of the models. Since natural data can continuously be collected the model parameters can be adjusted in a real-time manner in such a way that the model will adopt if user behavior in the field changes. In addition, the vast amount of data that can be collected through telematics would also enable the inclusion of multiple different parameters, such as contextual information about the driving situation, into the models.

Beta Testing After the realization phase (see Figure 2.1), the main goal is to implement the functionality of the developed features and to ensure a seamless integration into the vehicle environment. Currently, the automotive product development process is a pure stage-gate concept with fixed milestones, requiring full vehicle testing before a new feature can be deployed. However, UX professionals need more agile and data-driven development practices to implement UCD practices. The ability to run A/B experiments and obtain quantifiable data on user acceptance of a feature is essential to developing designs that meet customer needs. This allows designers to test new features, compare them with each other, learn how users respond to them, and optimize features already in use [232]. To enable Continuous Experimentation (CE) as it is already available in different digital domains [207, 233], several challenges need to be addressed [234]. Not only detailed usage data from production vehicles is needed, but also challenges related to the organizational and legal framework (see Section 5.2.1) need to be solved.

Continuous User Feedback The main advantage of implicit feedback is that it can be collected automatically and unobtrusively over a long period of time. This opens up many application areas for applications based on such data; from single driver behavior analysis to aggregated results of different user groups, from the short-term learning process to long-term UX. Currently, implicit feedback, collected in naturalistic driving studies, is mostly used for episodic UX analysis, such as evaluating a few months of driver behavior, conducting usability testing, behavioral hypothesis testing, and other activities. However, UX experts need cumulative UX assessment, recollecting different periods of use, such as the learning process, usage process or behavioral adaption over time. [57]. The vast amount of natural data that can be collected over the whole product lifecycle bears great potential to enable such analyses. Several studies indicate ongoing research in this direction. For example, Marrella and Catarci [235] propose implicit metrics for learnability evaluation, looking at deviations between the expected user behavior and actual user behavior, based on the analysis of usage data. This approach can quantify the degree of learnability over time and assists in identifying potential learning issues. In another study, Gerostathopoulos et al. [236] present the first attempt to use machine learning algorithms for automated learnability evaluation implementing automated quality gates.

Measurement of Subjective UX Factors To assess subjective UX factors such as trust, perceived safety, satisfaction, usefulness, acceptance, and others, qualitative methods, such as self-report methods are considered better suited than data-driven methods. Nevertheless, implicitly collected data can also be used to derive metrics for the validation of subjective UX measures. For example, in the web domain, Fox et al. [237] investigated which implicit metrics are correlated with user satisfaction to evaluate if explicit user satisfaction ratings and implicit user interest metrics could be cross-validated. Another example is presented by Lachner et al. [238]. The authors show that website visitors from different countries show significantly different usage patterns, suggesting that even personal characteristics, that influence the experience of a user, can be measured using quantitative metrics. Whereas being relatively unexplored, the measurement of UX based on implicit feedback could be a great advantage for the automotive and general UX design process [52].

Data-driven methods that leverage large amounts of interaction, glance, and driving data have great potential to improve current practices in the automotive UX design process. Applications based on machine learning algorithms are already successfully used in digital domains and have great potential for wider use in the automotive design process. The hypothetically large amount of natural data from production vehicles can provide UX experts with numerous applications to evaluate driver behavior and design even after product release. Based on this implicit feedback, the impact of design decisions can be quantified and their importance can be measured. This can improve and support decision-making and can make the UX design process more user-centric.

5.2.4 Recommended Actions to Better Integrate Data-Driven Methods into the UX Design Process

Having discussed the potential and limitations of data-driven, as well as the explicit needs of practitioners, we propose actions that can assist OEMs and practitioners to better integrate data-driven methods into the UX design process (RQ4). Although we the proposed actions do not guarantee completeness or success, we are confident that, based on the diverse experiences from the studies and close collaboration with the OEMs, they provide an important foundation for establishing data-driven UX as a practice in automotive development. The measures presented below are therefore intended to show the direction in which research should be conducted in order to bridge the gap between data-driven methods and the automotive UX design process.

Incorporate Data-Based Evidence in Decision-Making Processes Currently, most design decisions are made based on opinions or subjective assessments of individuals. In cases of disagreement, the results of small qualitative user studies are often overruled. We, therefore, argue that it is necessary to integrate policies or processes ensuring that each assumption that is made regarding the importance of a feature or its usefulness is backed up by statistical evidence. The provided statistical evidence can then be used to tailor qualitative studies to investigate the identified problem in detail. By doing this OEMs can not only be more confident that their product will meet the user's needs, but they can also save money that would have been spent for implementing or researching a feature that does not benefit customers in any way.

Increase Interdisciplinary Collaboration To fully exploit the potential of data-driven methods for the UX design process, it is important to merge the expertise of data scientists and UX experts. Data-driven evaluation methods should be developed in close cooperation with the UX experts so that they can easily access and interpret all relevant information. Only close collaboration between data scientists and UX experts can ensure that the need for intuitive data visualizations is met. Further, Yang et al. [213] argue that *"there is a real need for design tools and methodologies that support designers who lack constant access to capable data scientists"*. In addition, they present several best practices on how to incorporate machine learning into the design process. On the other hand, it is also important to empower UX experts by increasing their knowledge so that they can work with data or even leverage machine learning approaches. While there are several books and online courses to help designers learn about statistics and machine learning [239, 240], it seems that this knowledge is not yet so widespread in the automotive industry. The goal must be to provide UX experts with the necessary knowledge and tools so that basic statistical expertise is available. If automatically aggregated statistics are easily accessible to UX experts and product management, it will be less of a burden for UX experts and product managers to use statistical analysis to either make decisions or test hypotheses.

Introduce Clear Technical Specifications One of the biggest limitations automotive UX experts and data scientists complain about is the lack of specification and documentation. Therefore, we argue that any new feature that is developed must meet the interface specifications dictated by an overarching logging framework. This allows user behavior and interaction data to be analyzed for all features within the system. In addition, when a user-facing feature is being developed, UX experts need to be involved in the early stages of functional feature specification. They need to clearly articulate their requirements for how the feature should be evaluated and what data points are needed. This practice aims to avoid the common problem that specific signals needed to evaluate user behavior are not available due to insufficient specification in the early stages of product development.

Reduce Silo Mentality and Introduce Data Transparency One of the most frequently mentioned limitations is the lack of knowledge and documentation about what data is available, how to access it, and who is responsible for it. As a result, practitioners often do not even consider basing their decisions on data. One of the UX experts, interviewed in Study 3, explains that *"[...] we have to ask several people throughout the company to get the data. This slows us down because it can take a relatively long time until we get something useful. Most of the time we can't wait that long because we have to make progress with our designs"*. One way to counteract this is to introduce a centrally responsible unit that maintains an OEM-wide data catalog containing all available data points, their functional documentation, and current and/or intended use cases. In addition, this unit should also handle all legal approval processes for each signal. It is necessary to provide practitioners with clear guidance on what information is needed so that they are empowered to use data-driven methods in their daily work.

Introduce Agile Practices and Modernize Infrastructure One of the most discussed questions when it comes to automotive software development is how agile software development practices including Continuous Integration (CI) and CE can be integrated into the automotive development process. Hohl et al. [241] and Katumba and Knauss [242] describe several challenges OEMs face in their software development that are organizational and social in nature. These include long communication chains, low cross-functional mindset, high compliance and validation efforts, and technical challenges. While it is desirable to implement agile practices throughout the software development process, CE as an experiment-driven development approach is of particular interest for the UX design process. Many of the advantages of CE, which are well established in other application areas, can be transferred to the automotive industry [234]. However, to realize the potential of CE, several challenges such as security concerns or hardware-related resource constraints need to be addressed. While many studies focus on conceptual analyses regarding the use of CE in cyber-physical systems, only a few works present concrete solutions [209]. Giaimo et al. [209] propose a prototypical implementation and discuss design criteria to enable CE, but also note that their approach is not close to commercial use. It is, therefore, necessary to investigate how current challenges can be addressed and how CE practices can be put into practice so that software-based automotive designs can be evaluated similarly to web pages or mobile applications.

5.3 Discussion

Figure 5.3 summarizes our findings and relates them to each other. The figure shows the conflicts between the needs we collected in RQ2 and the limiting factors in the automotive domain (RQ1). In the following, we discuss some of these conflicts and relate them to the recommendations that emerged from our studies (RQ4).

The studies have shown that access to personalized or pseudonymized data is important for both the development and evaluation of intelligent functions in the car. Personalized or pseudonymized data is particularly important for customer research and evaluation tasks such as learnability assessment. However, qualitative methods, such as extensive user surveys or lab experiments, used to reduce uncertainty early in the design process are often costly. Here, the costs associated with user studies can be avoided with data-driven approaches based on customer usage data. However, each newly requested signal must go through an internal review process to ensure that it does not potentially contain personal information. OEM processes are lengthy and not clearly defined, resulting in delays even for data points that do not contain personal information. We do not see a technical solution to the conflict between the need for personalized or pseudonymized data and data protection regulations. However, the legal assessment can be supported by an early guideline that specifies how datapoints of new features must be documented and which legal requirements they must meet. In addition, by clearly defining and streamlining internal risk assessment processes, OEMs can also minimize the impact of such processes on non-personal data.

Another conflict arises between the need for data transparency and the current vehicle architecture, which consists of a large number of distributed subsystems. UX experts and data scientists need access to detailed data documentation from the various data sources in order to generate data-driven customer insights. However, because components are

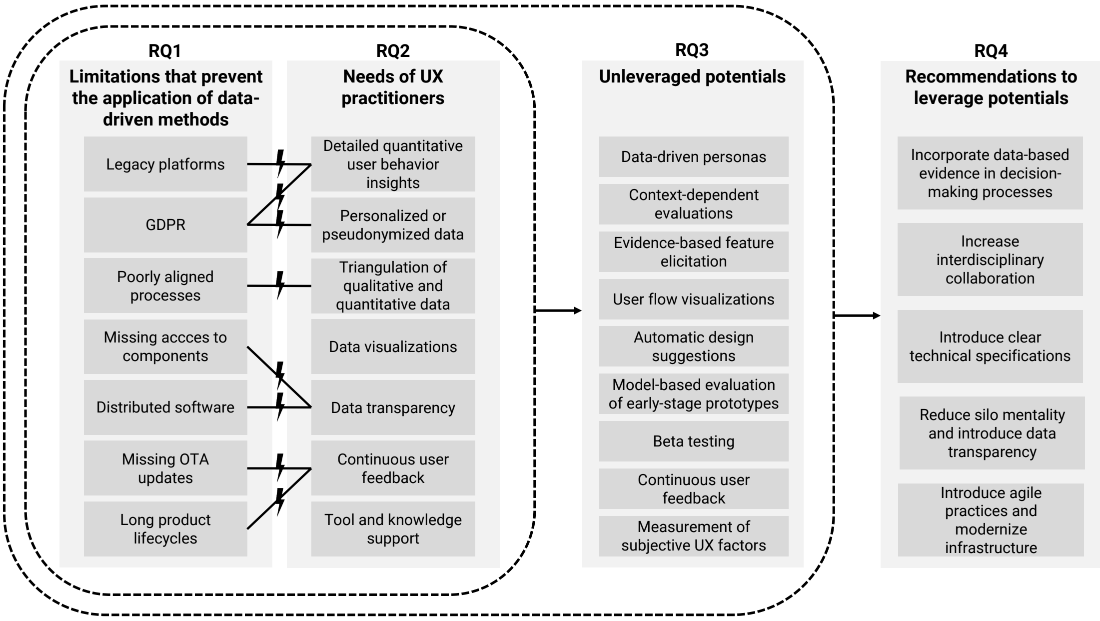


Figure 5.3: Summary of the results in accordance to Figure 5.1. The solid lines and lightning bolts indicate which specific limitation conflicts with which need. The dotted outlines and arrows indicate the consideration of combined previous results.

often developed by multiple suppliers, access to data points within these subsystems is typically limited. In addition, practitioners lack an overarching entity responsible for managing and interpreting individual data points in a holistic manner. As a result, a holistic picture of the available data points is often not available to the UX experts or can only be assembled with great effort and remaining uncertainties. However, several scenarios are conceivable to reduce the barriers and remaining uncertainties. From an organizational point of view, it is necessary to establish a central coordinating role in the development that provides a holistic overview of the available data points in the vehicle. This facilitates traceability and makes it easier to identify all the signals relevant to a particular problem. In addition, the silo mentality between different departments within the OEM needs to be broken down to promote interdisciplinary collaboration and efficiency. It should be noted, however, that this is an organizational-cultural problem that cannot be solved uniformly, nor does it apply equally to all OEMs.

Another related conflict exists between the need for data triangulation and poorly aligned processes when it comes to integrating data-based evidence into the UX design process. By combining quantitative and qualitative data, UX experts can explore and examine user behavior from different perspectives to gain a better understanding of the underlying problem. This can only be achieved if OEMs strengthen the interdisciplinary collaboration between data scientists and UX experts and make it mandatory to include data-based evidence when making design decisions.

The product development process in the automotive industry with its fixed milestones conflicts with the needs of UX experts to enable modern development methods such as CE. An iterative process for conceptualization and exploration is not explicitly defined in the automotive stage-gate model, which introduces potential for future research based on the recommendations presented in the main part of this thesis. Focusing on the early stages of product development, smooth integration of data in UX research activities helps product developers reduce uncertainty regarding potential customers and scenarios in which IVISs are used. We claim that the maturity of UX concepts can be improved with relatively little effort. Currently, existing potentials often cannot be exploited due to technical limitations. In addition, we found that the central goal of usage data collection is to satisfy management rather than to explicitly answer questions relevant to the UX design process. However, the requirements for eliciting natural interaction data should be initiated by UX experts and the problems they face in their daily work. The appropriate data points need to be defined based on the question posed by the UX experts.

In general, data-driven support is well anticipated throughout all UX design phases and can act as an enabler for multiple methods that bring the design and evaluation of IVISs to another level. However, it is not only the technical limitations specific to the automotive domain that conflict with the needs of UX experts and hinder the potential to be exploited. Insufficient transparency, specification and documentation of implicit vehicle data, lengthy processes, as well as a lack of integration of data-specific requirements in the early design phases lead to OEMs lagging behind their capabilities when it comes to data-driven and user-centered design of IVISs. The identified conflicts between practitioners' needs and current limitations and our initial recommendations serve as the basis for further research to develop organizational, technical, and legal solutions.

5.4 Conclusion

Based on a multiphase mixed methods approach that combines the results of four different studies, we elaborate on the needs, potentials, and limitations of data-driven methods in the automotive UX design process. By analyzing the problem at hand from different perspectives, we provide a first overview aimed at narrowing the gap between the automotive UX design process and data-driven development practices. UX experts articulate a clear desire for better integration of data-driven methods into the UX design process. To make the current design process more data-driven and thus more user-centered, UX experts need detailed user interaction data, tools and visualizations that make complex analysis results easily accessible, and methods that allow for the triangulation of qualitative and quantitative data. Furthermore, there is a strong need to integrate development processes such as CE, long used in web design, into the automotive UX design process. Our results show that approaches based on in-car data can improve the UX design process in many ways. Methods such as data-driven personas and feature usage analysis, which complement insights from traditional market research and qualitative studies, facilitate user-centered decision making. On the other hand, model-based design evaluations or context-dependent design suggestions can be seamlessly integrated into the design process itself. However, our results show that several conflicts need to be resolved in order to exploit the extracted potential and satisfy the needs of UX experts. Therefore, we recommend that automotive OEMs need to rethink their current decision-making process when it comes to feature and requirement elicitation. They should strive to consistently incorporate data-driven evidence into all design decisions that affect user-facing features. In addition, we argue that the technical requirements for capturing detailed user interaction data must be integrated into early product development processes. This requires strengthening interdisciplinary collaboration between data scientists and UX experts, transparently distributing relevant technical and legal information within the OEMs, and addressing the pervasive problem of silo mentality.

Acknowledgments

We would like to express our gratitude to the Swedish Innovation Agency VINNOVA for the project's funding (grant no 2018-05017). Also, we are particularly grateful to the industry practitioners who supported this study, sharing their knowledge and experience. Additionally, we would like to thank all the interviewees for their participation without whom the study would not have been possible.

Part II

Visualizing Driver Behavior

The results presented in Part I show that the design process of IVISs can benefit from the analysis of customer usage data throughout all design phase. However, designers often lack access to the data and the appropriate tools to analyze it. UX experts express the need for visualizations and tools that allow them to quickly and easily visualize usage data to understand how customers use their products (see Chapter 4). However, as described in Section 2.2.3, creating effective visual analytics tools is challenging and commercially available tools often do not meet the domain-specific needs of practitioners. Furthermore, visualizations need to be integrated into a larger workflow and should allow practitioners to explore data at multiple levels of granularity. In this chapter, we address these challenges. In Chapter 6, we propose and evaluate a multilevel user behavior visualization framework for touchscreen-based IVISs, consisting of three visualizations on different levels of granularity. In Chapter 7 we further refine these visualizations and integrate them into an interactive visualization tool called ICEBOAT. ICEBOAT visualizes driver interactions and driving behavior on different levels of detail, allowing easy comparison of user flows based on performance and safety metrics. We evaluate ICEBOAT with 12 UX professionals and show that it facilitates their decision-making process.

Chapter 6

Visualizing User Interactions with IVISs

Context The studies presented in Chapter 4 and Chapter 5 reveal the lack of data-driven methods and big data analytics when it comes to analyzing driver interactions with IVISs. While OEMs already collect a large amount of data from their vehicle fleets, it's not being used to its full potential. This is especially true for the design and evaluation of IVISs. In the absence of evidence-based customer insights, decisions at all stages of UX design must often be made intuitively. Our previous studies, presented in Chapter 4 and Chapter 5, show that automotive-specific aggregation and visualization methods that allow UX experts to independently explore customer behavior could effectively address this problem. UX experts need tailored aggregation and visualization methods to gain insight into user and driving behavior. They express the need for methods that visualize not only the user interactions, but also the glance behavior and the contextual factors that describe the driving situation.

Contribution We propose a *Multi-Level User Behavior Visualization Framework* for touchscreen-based IVISs consisting of three different levels of abstraction: (1) The task level that visualizes alternative interaction flows for one task (e.g., starting navigation), (2) the flow level that visualizes metrics of interest for the different interaction sequences of one flow (e.g., using the keyboard vs. using Point of Interests (POIs) to start navigation), and (3) the sequence level that augments single interaction sequences with contextual driving data such as speed or steering angle. UX experts can use the visualizations to effectively gain insights into user flows, their temporal differences, and the relation between user interactions, glance behavior, and driving behavior. The presented visualizations were found very useful in an informal evaluation study with 4 automotive UX experts.

Related Publications This chapter is adapted with minor changes from Ebel et al. [3].

6.1 Multi-Level User Behavior Visualizations

Automotive UX experts need visualization methods that allow them to analyze how drivers interact with the IVIS in order to generate customer insights and make user-centered design decisions. (see Chapter 4). Our approach supports a holistic user behavior evaluation of the IVISs by visualizing different levels of abstraction of the driver's interaction with the center stack touchscreen interface. We have designed these three different visualizations as each of them satisfies specific needs introduced in Chapter 4. The Task Level View allows UX experts to explore how users navigate the system, what the main interaction flows are, and how they relate to each other. The Flow Level View provides a quantitative comparison of different flows based on a chosen metric. Finally, the Sequence Level View allows UX experts to analyze specific sequences in terms of the relationship between touch interactions, glance behavior, and driving context.

The data used in this work is collected from production vehicles without a specifically designed test environment or a defined group of participants. This theoretically allows the data to be collected from any modern car in the fleet of our Mercedes-Benz, a leading German OEM. The use of natural data has three main advantages over data obtained from laboratory experiments: (1) a large amount of data can be collected from the entire user base; (2) there are no specific costs for controlled experiments; (3) the context of use, i.e. the driving situation, is inherent in the data. In the following, the data collection and processing framework is introduced, followed by a detailed description of the visualizations mentioned above.

6.1.1 Data Collection and Processing

The visualizations shown in this chapter are based on data from 27,787 trips of 493 individual test vehicles collected through the Telematics Data Logging Framework. The vehicles are used for a diverse range of internal testing procedures. No special selection criteria were applied and therefore all vehicles with the most recent telematic architecture contributed to the data collection. The event sequence data consists of timestamped events containing the name of the interactive UI element triggered by the user and the type of gesture detected. The extraction of interaction sequences differs slightly from the procedure presented in Section 3.2.1, as described below. First, all event sequences that satisfy the start and end condition (e.g., the respective UI elements) of a task and do not meet a task-specific termination criterion are extracted and are assigned a *Task ID*. The termination criterion is intended to give users the ability to customize the evaluations to meet their needs. It can be defined as a set of specific UI elements or a maximum time limit t_{max} that applies to the interval between two interactions (similar to Δt_{max} as introduced Section 3.2.1). All sequences in which the termination criterion is met are cleansed. If, for example, it is defined that a maximum of 60 seconds t_{max} = 60 may elapse between two events and otherwise the task is considered incomplete, all sequences in which this applies are cleansed. After sequence extraction, the data processing procedure is the same as described in Section 3.2.1.

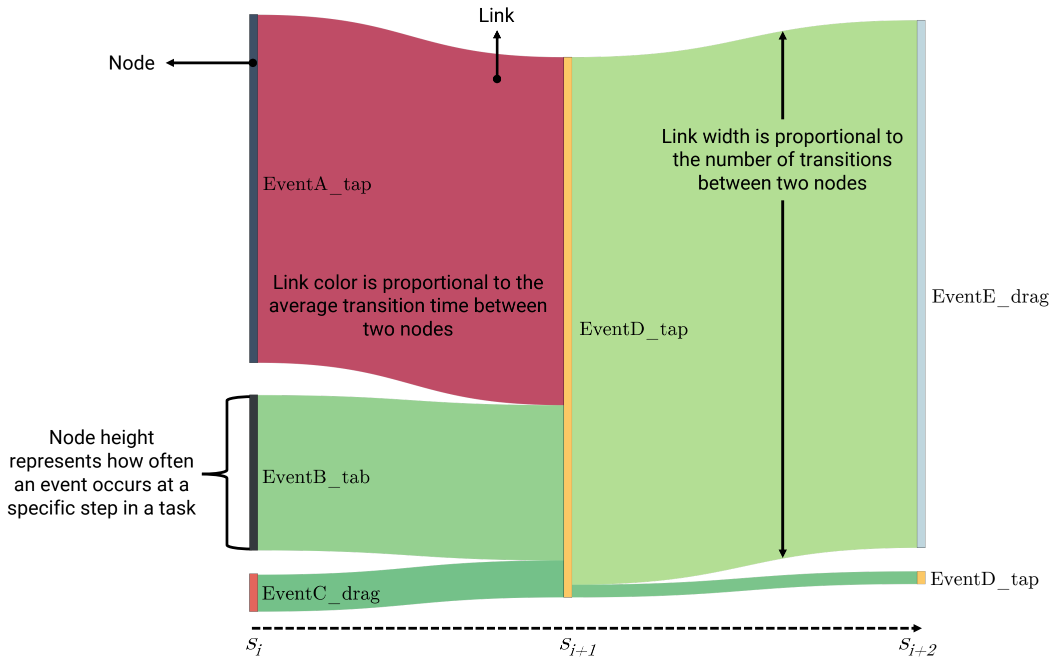


Figure 6.1: Visual encoding of nodes and links in the Task Level View.

6.1.2 Task Level View

The Task Level View visualizes how users navigate through the system to accomplish a specific task. Event sequence data generated by touchscreen interactions is aggregated and visualized in the form of an adapted Sankey diagram. We chose Sankey diagrams as the basis for the Task Level View because of their popularity and efficient way to visualize multiple different flows and their distribution. We address the main weakness of Sankey diagrams, which is that they do not encode temporal information, by introducing color-coded links. Being able to see the most common user flows and their temporal attributes at a glance helps UX experts to identify unintended or unexpected user behavior. The individual components and their visual encoding are shown in Figure 6.1.

Nodes Each node represents an event at a particular step in a task. The nodes are visualized as rectangles whose height is proportional to the cardinality of the event at a given step in the task. The name of the UI element and the gesture (annotated as *_tap, _drag or _other*) used for an interaction are displayed next to the Node (see Figure 6.1). The horizontal position indicates the step in the flow at which the event occurred. Thus, nodes that are vertically aligned represent events at the same step in a task. In Figure 6.1, step s_i contains three different events, while s_{i+1} contains only one event, meaning that whatever users did in s_1, they all made the same interaction (*EventD_tap*) in s_{i+1}. Nodes representing the same event at different steps are colored the same (compare *EventD_tap* in Figure 6.1). Hovering over a node displays the number of entities, incoming and outgoing links.

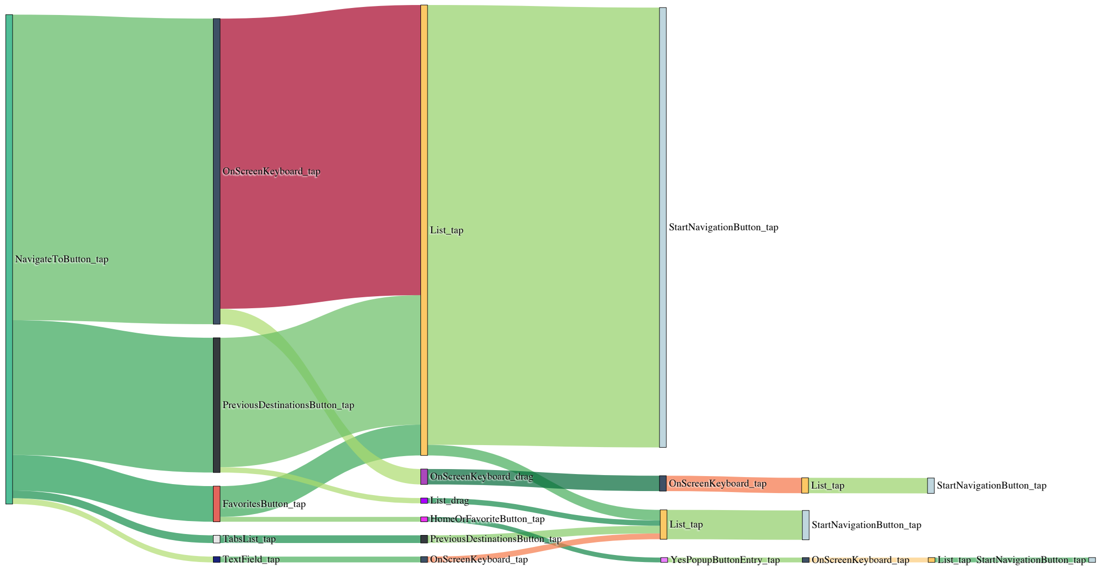


Figure 6.2: Task Level View where ($t_{max} = 60s$ and $p_{min} = 0.005$).

Links Each link connects two nodes and therefore represents a transition between two events. The link width is proportional to the number of transitions between the source node and the target node. The link color represents the average transition time between two events. The time is normalized to [0,1] using min-max-normalization, with higher values representing slower transitions. The normalized values are mapped to a linear color scale from green (0; short time) to red (1; long time). As displayed in Figure 6.1, the transition $EventA\_tab \rightarrow EventD\_tab$ is the most prominent one moving from s_i to s_{i+1} but also the slowest. When hovering over a link, additional information is given describing in how many sequences (absolute and relative values) users went from the source node to the target node and how much time it took on average.

To create a visualization, the events that indicate the start and the end of a task need to be defined. The optional parameter p_{min} allows users to set a lower bound, such that only flows with a relative frequency greater than p_{min} are displayed. This filter increases readability since Sankey diagrams are hard to read for a large number of nodes [93]. Additionally, UX experts can define a set of interactions that are represented as a single node even if they occur multiple times in succession (e.g., keyboard taps).

Example Figure 6.2 shows the Task Level View for a navigation task that starts with opening the navigation app from the map view on the Homescreen (*NavigateToButton_tap*) and ends with confirming that the route guidance shall be started (*StartNavigateButton_tap*). Investigating the different flows, one can clearly see that, whereas in most of the cases users directly started to use the keyboard to enter their destination (62 %), some users chose to use the option to select a destination out of their previous destinations (28 %) or their pre-entered favorites (7 %). After typing on the keyboard (*OnScreenKeyboard_tap*) to enter the destination, the majority of users directly chose an element out of the list of suggested destinations presented by the system (*List_tap*). Afterward, the majority then started the route guidance by accepting the proposed route

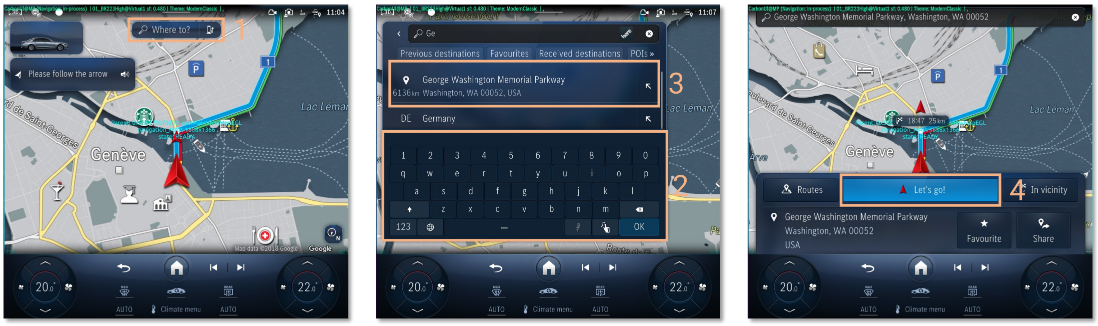


Figure 6.3: Sample user flow of a navigation task. Flow 1: (1) *NavigateToButton_ tap* → (2) *OnScreenKeyboard_ tap* → (3) *List_ tap* → (4) *StartNavigationButton_ tap*.

(*StartNavigateButton_ tap*). An example of this flow and how it looks like in the production vehicles IVIS is given in Figure 6.3. Apart from identifying the most popular flows, the Task Level View also assists UX experts in finding unintended user behavior. For example, after the first interaction (*NavigateToButton_ tap*) the keyboard automatically opens and users can directly start typing. However, roughly one percent of the users first clicked on the text field and the started typing. This could lead to the hypothesis that users did not anticipate that the text field is already pre-selected and that they therefore tried to activate it by clicking on it.

Apart from visualizing certain user flows and their popularity, the color-coding of the links allows conclusions to be drawn about interaction times. Typing on the keyboard (*OnScreenKeyboard_ tap*) is by far the most time-consuming interaction in the presented task. Since it is the only aggregated event consisting of multiple user interactions this information may not be surprising. It nevertheless shows that a large portion of the time on task can be attributed to typing on the keyboard. Taking a closer look at the second step of the task, one can observe that users need about 2.3 seconds to choose a destination out of a list of pre-entered favorites (*FavoritesButton_ tap*) is, whereas they need roughly 3 seconds to choose a destination out of a list containing all previous destinations (*PreviousDestinationsButton_ tap*). This difference could be attributed to the fact that the favorites list is a structured list that tends to have fewer entries than the chronologically sorted list of previous destinations.

6.1.3 Flow Level View

Whereas the Task Level View provides an overview of the different flows and their proportion, other metrics like for example the time on task of specific flows and how they compare are not sufficiently visualized. The Flow Level View (Figure 6.4) addresses this shortcoming by visualizing the distribution of a certain metric (for the example we use the time on task) of all sequences that belong to a flow (see Figure 6.4). By visualizing the time on task as violin plots, two main insights can be generated. On the one hand, multiple statistics (e.g., min/max, mean, interquartile range) are visualized when hovering over the plot. UX experts can assess the displayed metrics and compare them to target values or industry guidelines [158, 243]. On the other hand, displaying the violin plots next to each other allows a visual comparison of the individual flows. For example when comparing the distribution of flow 1 (*NavigateToButton_ tap* → *OnScreenKeyboard_ tap* → *List_ tap* → *Start-*

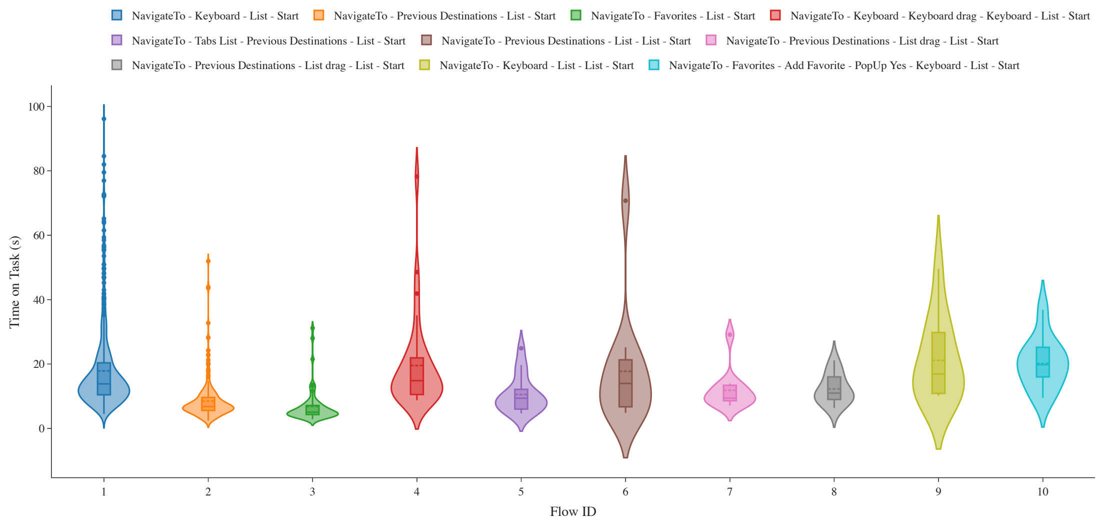


Figure 6.4: Flow Level View with shortened event names.

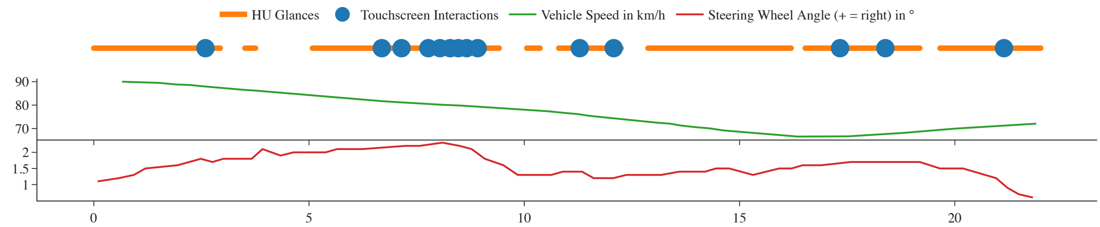
NavigationButton_ tap), flow 2 (*NavigateToButton_ tap* → *PreviousDestinationsButton_ - tap* → *List_ tap* → *StartNavigationButton_ tap*), and flow 3 (*NavigateToButton_ tap* → *FavoritesButton_ tap* → *List_ tap* → *StartNavigationButton_ tap*) one can observe that the time on task when using the keyboard is nearly double the time needed compared to either using the favorite or previous destination options. Comparing the latter (flow 2 and flow 3), using the favorites option is about two seconds faster than using the previous destination option. Whereas this difference has already been identified in the example describing the Task Level View, the impact on the whole task completion time can now be quantified.

6.1.4 Sequence Level View

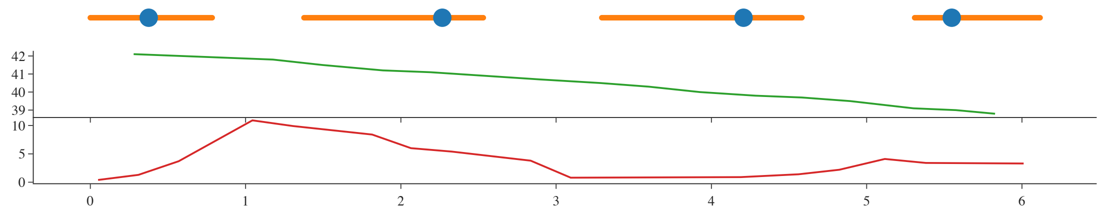
An increased visual distraction from the driving task toward non-driving-related tasks is associated with increased crash risk [122]. Thus, insights into the interrelation of user interactions, glance behavior, and driving behavior can yield valuable information for UX experts regarding the safety assessment of touchscreen-based IVISs. Whereas the previous views visualize general trends, the proposed Sequence Level View (see Figure 6.5) generates such insights by making it easy to identify long off-road glances, demanding click patterns, or other safety-critical driving behavior.

The visualization consists of two main parts: The upper part is an overlay of touchscreen interactions (blue dots) and the driver's glances toward the center display (orange lines). Each dot represents one interaction and each line indicates the duration of a glance toward the display. The lower visualization, consisting of two graphs, represents the driving-related data (vehicle speed (green line) steering wheel angle (red line)). In Figure 6.5, three different sequences are visualized, emphasizing the importance to set the evaluation of user flows in perspective to the context.

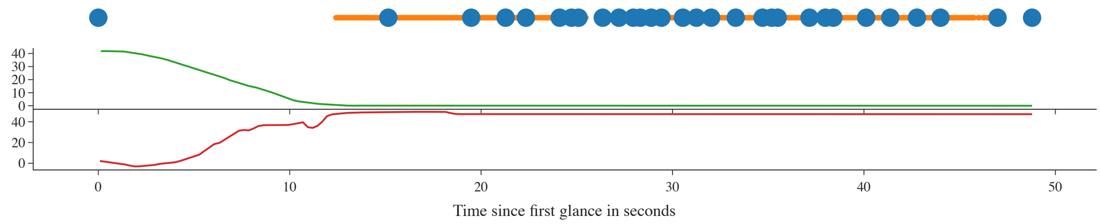
In Figure 6.5a a specific sequence of Flow 8 is visualized. One can observe that it took the driver five long glances ($t > 2\,\mathrm{s}$) and three short glances ($t < 2\,\mathrm{s}$) to fulfill a task of 14 interactions whereas 10 of the interactions are keyboard interactions. Additionally, we can observe that the vehicle speed decreased after starting to type on the display



(a) Flow 8: *NavigateToButton_tap* → *OnScreenKeyboard_tap* → *List_drag* → *List_tap* → *StartNavigationButton_tap*



(b) Flow 2: *NavigateToButton_tap* → *PreviousDestinationsButton_tap* → *List_tap* → *StartNavigationButton_tap*



(c) Flow 6: *NavigateToButton_tap* → *TextField_tap* → *OnScreenKeyboard_tap* → *List_tap* → *StartNavigationButton_tap*

Figure 6.5: Sequence Level View showing three individual secondary task engagements.

and increased again at the end of the sequence. The change in the steering wheel angle is generally low, however, one can detect a small drift during the first intense typing interaction and a small correction after the second long glance. Whereas the first sequence took around 20 seconds for completion, the sequence using the previous destination option only took roughly six seconds, requiring four glances and four interactions. The vehicle speed did only slightly decrease during the interaction. In contrast to the two above sequences, the sequence displayed in Figure 6.5c consists of 30 touch interactions (25 keyboard interactions) but only two glances. During normal driving, taking the eyes off the road for such a long period of time would be considered highly safety-critical. However, considering the vehicle speed and the steering wheel angle, one can conclude that the driver pulled over to the right and stopped the car before starting to interact with the HMI. Therefore, this is not considered critical behavior and shows that certain statistical outliers need to be assessed individually.

6.2 Informal Evaluation

To assess the usefulness of the proposed approach and to answer the question of whether the visualizations are suited to generate knowledge from large amounts of event sequence data, we conducted a user study. The goal of the study was to understand how participants interact with the presented visualizations when trying to answer questions regarding user behavior. Therefore, we recruited four automotive UX experts (P1-P4, one UX Researcher, and three UX Designers with 3, 9, 4, and 18 years of working experience respectively). Two participants were directly involved in the design and development of the HMI analyzed in this study. The examples presented in the previous sections were sent to the participants as an interactive web page and a document containing further information regarding the presented interface was provided. Due to the Covid-19 pandemic, we conducted the interviews remotely using Zoom. During the study, the participants were asked to share their screen and the interviews were recorded using the built-in audio and video recorder. Each interview comprised an introduction (20 minutes), an interactive part (30 minutes), and a discussion (10 minutes). During the introduction, we presented the objective of the presented system, the telematics framework, the exemplary task (screenshots and the respective UI elements), and demonstrated the features of the system. We asked the participants to explore the different visualizations and to ask questions in case some explanations were unclear. During the interactive part, the participants were asked to answer a list of seven distinct study questions (see Table 6.1). The questions are inspired by the needs and potentials identified in Chapter 4 and aim to test whether the visualizations are suitable for generating the expected insights.

After interacting with the visualizations to answer the study questions, the participants were given another 10 minutes to explore the visualizations to find any behavioral patterns that might indicate usability issues. After the interactive part, we initiated a discussion regarding the different visualizations and how the participants might integrate them into their design process. After the interview, the participants were asked to answer a survey[1] with 8 questions addressing the usefulness of the system and its potentials with regard to their workflow. The questions demanded answers on a 7-point Likert scale ranging from strongly disagree (1) to strongly agree (7).

6.2.1 Generated User Behavior Insights

In the following, we assess whether the visualizations are suited to answer the study questions (see Table 6.1).

Task Level View All four participants answered the questions regarding the Task Level View (SQ1-SQ3) without additional support. They compared the respective links and nodes to answer SQ1 and SQ2 and interpreted the color coding as intended to find the most time-consuming interaction (SQ3). Also, P1 and P2 were particularly interested in flow 4: “I can easily see that most people use our system as intended and I’m not overly concerned with flows that only occur very few times. But seeing 5 percent using the drag gesture on a keyboard [...] I would like to get into more detail” (P1). Interestingly, during the interviews, we observed that the participants used the Task Level View as a

[1] Questions and results are given here: `http://kups.ub.uni-koeln.de/id/eprint/65348`

Table 6.1: Study questions and objectives.

#	View Level	Question/ *Objective*
SQ1	Task	Which path do most users take to start the navigation? *Traverse graph and interpret link width*
SQ2	Task	Do users prefer the favorites or previous destination option? *Interpret node height*
SQ3	Task	Which interaction is the most time-consuming? *Interpret link color*
SQ4	Flow	What is the fastest way to start the navigation? *Interpret metrics shown as hovering elements*
SQ5	Flow	Which flows are interesting to compare and why? *Compare distributions to find distinctive features*
SQ6	Sequence	Can you observe any safety-critical behavior? *Interpret glance duration and click behavior*
SQ7	Sequence	How do you interpret the driving situation? *Interpret driving parameters*

kind of reference. Often when an anomaly or a pattern of interest was detected in one of the other views, participants invoked the Task Level View to verify what role the flow or the specific interaction plays in the overall context of the task.

Flow Level View Compared to the Task Level View, only two participants (P1 and P3) answered SQ4 without any further information. They quickly decided to base their answer on the median time on task and therefore identified flow 3 to be the fastest way to start the navigation. Whereas P1 and P3 were familiar with boxplots and violin plots, this kind of visualization was unknown for P2 and P4. P4 stated that *"[he] would need to get more familiar with this kind of statistics"*. They, therefore, needed some additional assistance, but then solved SQ4 in similar a manner as P1 and P3 did. P1 adds that: *"Interpreting this visualization gets easier the more often one uses it in the daily work"*. When asked to compare flows that might yield interesting insights, P1 argued that the distribution of the time on task could be used as a complexity measure that a more widespread distribution could indicate a more complex flow. Therefore, the interviewee compared flow 1 and flow 6, with the only difference being that in flow 6 people clicked in the text field before they started typing. Based on the more widespread distribution of flow 6, P1 argued that *"some people seem to have difficulties in understanding that the text field is already activated and that there is no need to tap on it. This seems to lead to longer interaction times"*.

Sequence Level View Working with the three different examples of the Sequence Level View (SQ6 and SQ7) all participants were able to derive certain hypotheses regarding driver distraction based on the glance and driving behavior. All participants found that the glances in Figure 6.5a are critically long. Regarding the long glance without any interaction after typing on the keyboard, P4 states that it *"[...] might be due to a slow internet connection or because the intended destination was not in the list of suggestions"*. Based on the vehicle speed and the steering wheel angle participants concluded that the person was distracted by the interaction and the long glances. P1 explains that *"[d]uring the keyboard interaction, there is an increasing deviation in the steering angle and a* *correction at the end of the interaction, even though it may be small in absolute terms"*. In contrast to Figure 6.5a, the glances in Figure 6.5b were considered not critical by all participants. P1 remarks *"[t]hat's one glance per interaction, just like we want it to be"* and further explains that one cannot attribute the deviation of the steering angle to the interaction with the HU. P3 was particularly interested in why people are in need to focus on the head unit after interacting with it and suspects that users want to have visual feedback on their interaction. Regarding the sequence visualized in Figure 6.5c all participants quickly identified that the driver pulled over to the right and then started engaging with the display. Therefore, they considered this behavior as not safety-critical.

6.2.2 Benefits and Use Cases

In general, participants agree, that the presented visualizations would benefit multiple use cases in the UX design process. Participants' statements describe that the three visualizations have great value for efficiently visualizing large amounts of interaction data and that they currently miss such possibilities in their daily work. P3 concludes that *"[a]ll the information that brings you closer to the context of the user while you are sitting in the office behind your screen is extremely valuable"*.

Task Level View The Task Level View is considered very useful by all participants. They, in particular, appreciated the simple and intuitive representation of user flows. This is also shown insofar as they had no problems answering Study Questions SQ1-SQ3. P3 was especially interested in flows that can be considered conspicuous because *"[y]ou can find issues where nobody would even think of doing a qualitative study because you did not even think of this behavior. But if 5 % of all people behave that way there must be a reason for it and it should be further investigated"*. P3 further added that *"[...] there could be so many feature improvements based on the issues detected using this view"*. Similarly, P2 adds that *"[they] currently have a collection of questions from different UX designers within the company that could, probably, be answered with this kind of visualization"*. The interviewee further describes that a data-driven platform similar to the proposed one could have great benefit not only for UX experts but also for management and product development.

Flow Level View In general, the participants agree that the Flow Level View is helpful in the design process. P1 states that *"[b]eing able to see statistics like the median and the distribution of the sequences makes this visualization valuable when comparing different flows"*. P4 argues that it would also be interesting to see how these graphs change over time when people get more familiar with the system: *"How do these graphs look like for novice users and how do they look like for experts users?"*. Furthermore, P1 adds that this would benefit the assessment of intuitiveness and learnability. P3 states that the distribution of sequences over the time on task is from particular interest because *"[...] if a lot of users are at the far end of the distribution it would mean that a lot of them might have problems with this flow and I would be interested in why it takes them such a long time to complete the task"*. P1 further elaborates that it would be helpful to see specific sequences for identified outliers since the time on task alone indicates critical behavior.

Sequence Level View All participants consider the Sequence Level View very helpful and argue that it plays an important role, especially in combination with the other

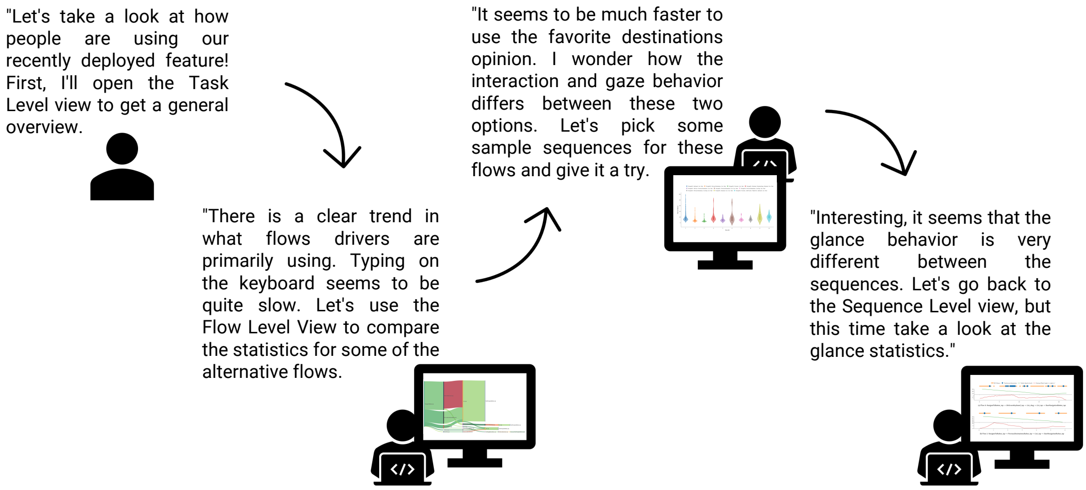


Figure 6.6: A hypothetical use case that illustrates our vision of how the visualizations can be used in the automotive UX design process.

visualizations. Whereas the other views present higher-level aggregated statistics, the visualization of specific sequences was helpful to develop a more precise understanding of how the interactions in the vehicle take place. The additionally given information and especially the glance behavior data was considered very useful because *"[o]ne can derive important information regarding the context to set the interaction into perspective"* (P4). Additionally, P3 emphasizes the importance regarding safety assessments because *"it might be better to prioritize something slower but with fewer glances"*. P1 and P4 both explain that in order to get insights into glance behavior they, until now, had to set up specific lab studies.

6.3 Discussion

The conducted user study shows that the presented visualizations help UX experts in designing IVISs, assisting them in finding usability issues and unexpected user behavior. An example that visualizes how the different views support each other and how UX experts may use them is given in Figure 6.6.They report that they would use performance data more often if such visualizations would be available and argue that the generated insights would benefit the feature and requirements elicitation process. The Task Level View was considered the most helpful, closely followed by the Sequence Level View, followed by the Flow Level View. This coincides with the observations made during the evaluation study. During the study, participants switched between the different views depending on the type of information they were interested in. This consolidates our assumption that the different views support each other in a meaningful way and that different levels of detail are necessary to generate the best possible insights into driver IVIS interaction.

Our results show that visualizing large amounts of automotive interaction data using the proposed three visualizations is promising. However, we also identified points for improvement. One common suggestion is the mapping between user interactions and actual screens. This helps to interpret the visualizations without the need to know the names of the UI elements. Additionally, participants suggested making the visualizations visually more pleasing and proposed adding a dashboard-like overview of general statistics. This being a first exploratory approach, we only evaluated if participants interacted as intended and if they were able to generate the anticipated insights. For future iterations, it would be interesting to assess effectiveness and efficiency and compare multiple alternatives. Additionally, future evaluations should include participants outside of Mercedes-Benz. None of our participants were affected by color vision deficiency, however, we have been advised to use a colorblind-friendly palette in future versions.

Even if they do not directly influence the contribution of this work, ethical aspects of data collection, data security, and privacy are particularly important in the broader scope of this work. As of now, only company-internal testing vehicles contribute to the data collection. However, for future use cases, it is conceivable that customers contribute to the data collection and receive benefits such as earlier access to new features as compensation. The consent for data collection is given actively using the so-called *"opt-in"* standard. Therefore, users have full control over the decision whether or not to share their data to contribute to product improvement. As already mentioned, the data is completely anonymized, making it impossible to draw conclusions about individual users or their behavior.

6.4 Conclusion

In this chapter, we present the Multi-Level User Behavior Visualization Framework, which provides insights into driver-system interactions with IVISs on three levels of granularity. The proposed approach leverages the Telematics Data Logging Framework (see Section 3.1), which collects live data from production vehicles. The presented visualizations are based on event sequence data, driving data, and glance behavior data. As a whole, they enable UX experts to quickly identify potential usability problems, quantify them, and examine their influence on glance or driving behavior using representative examples.

By addressing the data visualization needs of automotive UX experts as presented in Chapter 4, the proposed approach is a first step towards better integration of quantitative user behavior data into the automotive UX design process. We envision the presented approach to be integrated into an overarching analytics solution that allows UX experts to freely explore large amounts of live data collected from production or test vehicles to gain immediate insights into in-car user behavior.

Chapter 7

An Interactive User Behavior Analysis Tool for Automotive User Interfaces

Context Automotive UX experts can only make data-driven design decisions if they have the tools to effectively and efficiently visualize and evaluate drivers' interactions with IVISs. Such analytical tools need to be developed according to the needs of domain experts, must communicate results through meaningful visualizations [83], and should keep the overhead for users low [86]. In Chapter 6 we presented a Multi-Level User Behavior Visualization Framework that visualizes driver behavior data on three different levels of granularity. Our informal evaluation shows that the proposed visualizations can help UX experts to evaluate IVISs and find usability problems and unexpected user behavior. However, to support UX experts in evaluating IVISs, these visualizations need to be extended, connected, and embedded in an interactive tool that automates data processing and visualization generation.

Contribution In this chapter we propose ICEBOAT, an interactive visualization tool that enables automotive UX experts to effectively and efficiently analyze driver interactions with the center stack touchscreen to evaluate touchscreen-based IVISs. Following a mixed methods UCD approach, we conducted an interview study (N=4) to extract the domain specific information and interaction needs of automotive UX experts and used a co-design approach (N=4) to develop an interactive analysis tool. The tool visualizes user interaction data, driving data, and glance data, that is collected via the Telematics Data Logging Framework (see Section 3.1). UX experts can specify any task they want to analyze, either by manually defining the customer journey, or by using an interactive IVIS emulator. ICEBOAT automatically processes the data and generates various statistics and visualizations that are based on the visualizations presented in Chapter 6. An interactive drill-down concept allows UX experts to start wide and zoom in to analyze individual touchscreen interactions. UX experts can compare different flows according to performance-related and distraction-related metrics such as time-on-task, number of glances, or total glance duration. Our evaluation (N=12) shows that ICEBOAT enables UX experts to efficiently generate knowledge that facilitates data-driven design decisions.

Related Publications This chapter is adapted with minor changes from Ebel et al. [5, 8].

7.1 Approach

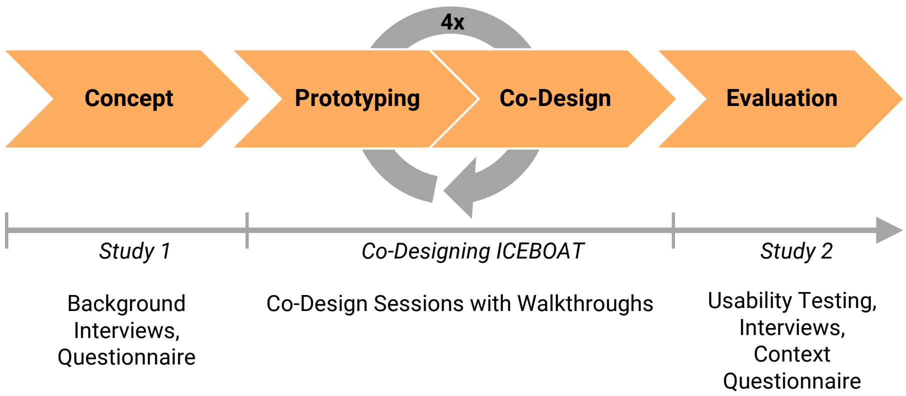


Figure 7.1: Summary of the mixed methods UCD approach.

The growing number of features of modern touchscreen-based IVISs and the need to evaluate them with respect to the driving context [11] makes it increasingly complex to design IVISs that meet user needs. This becomes evident when considering that the usability of infotainment systems has been the biggest source of problems for new car owners for several years [35, 36, 37]. As shown in Chapter 4, UX experts are often forced to neglect design evaluation due to time constraints or data accessibility issues. With ICEBOAT we introduce a tool that counteracts current shortcomings by enabling UX experts to effectively and efficiently visualize and evaluate drivers' interactions with IVISs.

To develop a solution that meets the needs of automotive UX experts and improves the design and evaluation process, we followed a mixed methods UCD approach [244, 245] (see Figure 7.1). In the first phase, we conducted semi-structured background interviews (n=4) to extract requirements according to information and interaction needs and to evaluate the visualizations originally proposed in Chapter 6. Using a participatory design approach, we then co-designed prototypes with four automotive UX experts. Throughout the co-design approach, we walked different focus groups through the current state of the prototypes and discussed potential improvements and necessary changes. After the fourth co-design session, we evaluated the prototype in a second study where we conducted usability testing and explored the context of use of the prototype.

7.2 Study 1: Extracting Information and Interaction Needs

The first study has two goals: First, we need to confirm the results of our informal evaluation study presented in Chapter 6. Second, to generate knowledge on how to extend and connect these visualizations to support UX experts in evaluating IVISs, we need to extract detailed information and interaction needs. For this purpose we conducted semi-structured interviews. The subsequent co-design process is based on the results of this interview study.

7.2.1 Participants

We conducted semi-structured interviews with 4 UX experts (I1 – I4). All of them had between five and nine years of professional experience. At least five of those years were spent in the field of UI design. They have also all been with Mercedes-Benz for more than five years. We therefore consider them to be knowledgeable about automotive design processes and working methods.

7.2.2 Procedure

The interview agenda consisted of three parts: *introduction*, *main section*, and a *conclusion*. Following the recommendations of Renner and Jacob [246], we prepared a set of open-ended questions and optional follow-up questions for each section. The latter were intended to refine ambiguous responses and further guide the interview. In addition, we periodically summarized the responses during the interview to reflect and confirm correct understanding. While the introduction was designed to create an open atmosphere and establish a common ground, we ended each interview with a conclusion, asking the interviewees if there was anything they wished to add. The main section contained the majority of the questions. Here we asked the interviewees about their *information needs*, *interaction needs*, and the *visualizations* they would expect to see in a potential tool that supports their current workflow. Regarding the visualizations, we gave the interviewees some time to brainstorm and develop their ideas. We then presented the interviewees with the visualizations presented in Chapter 6. By showing them the existing visualizations, we hoped to support their ideation process [244] and wanted to confirm the results of our informal evaluation study (see Section 6.2).

7.2.3 Information Needs

After analyzing the coded interview transcripts, we identified 39 information needs that fall into 7 categories. We present these needs below, where *Count* refers to the number of unique needs within a category and *Support* refers to the number of total needs expressed by the participants.

INF-1: Usability and Distraction-Related Metrics (*Count = 8, Support = 12*). Driver interactions with IVISs while driving are considered secondary tasks [121]. Thus, not only usability but also driver distraction play a major role in evaluating automotive user interface [2]. Accordingly, the respondents formulated various information needs that revolve around the understandability of the UI (I1, I2, I3), performance-related metrics such as time on task (I1), error rates (I1, I2), or the number of interactions needed to perform a task (I2). They also stated that for a holistic evaluation they need to be able to evaluate the visual demand (e.g. number of glances) of features (I1, I2) and individual user flows (I1). They also stressed the importance of being able to see the correlation between IVIS usage and driving data.

INF-2: Feature Usage Information (*Count = 8, Support = 12*). The automotive industry is moving from a technology-driven development approach to a more user-centric one [31]. While this process has been going on for many years, there is still a lack of knowledge about how features are used by customers. This leads to many features being carried over from old releases that may not be needed by customers (see Chapter 4). During the interviews, feature usage information was the first KPI that respondents thought of. The typical questions UX experts want to answer based on data insights are questions like *"How often is a feature used?"* (I1-I4) or *"How long is a feature used on average?"* (I1, I2). Participants indicate that information about feature usage is valuable because it is often used to decide whether to continue or discontinue a feature.

INF-3: Usage Pattern Visualizations (*Count = 7, Support = 11*). To gain deeper insights into user behavior, UX experts expressed different needs regarding the analysis of user flows and how users interact within certain features (*Usage Patterns*). They want to know how people use the system (I2, I4), how they navigate the system to perform certain tasks (I1, I4), and what kind of UI elements they use (I2). Participants are also interested in merging this information with usability and distraction-related metrics (e.g., to compare different flows).

INF-4: System Information (*Count = 6, Support = 10*). The cars in an OEM's fleet are very heterogeneous, both in terms of hardware and software. Not only do manufacturers offer different models that differ according to the market in which they are sold, but customers can also configure their cars according to their personal preferences (e.g., different sizes of center stack touchscreens) [247]. This, combined with the long product lifecycle and limited ability to perform OTA updates, especially for older models, results in many different UI versions being used by customers. This is reflected in the information needs of UX experts. They state that they need to compare usability and distraction-related metrics, feature usage information, and usage patterns across car models (I1-I4), software versions (I2, I4), screens (driver vs. front passenger vs. rear passengers), and screen sizes (I2). This information is needed to assess the interplay between hardware and software but also to track progress.

INF-5: Contextual Information (*Count = 5, Support = 6*). Driver behavior and driver interactions are highly context sensitive (see Chapter 8) and participants state that they need contextual information to better judge individual interaction sequences. For example, they state that they need information about the driving situation (1, 3) to be able to judge how drivers interact in different situations. They also want to know how many passengers were present (2) and whether a cell phone was connected to the IVIS (2), arguing that these could be additional sources that influence driver behavior without being represented in the interaction, glance, or driving data.

INF-6: Input Modalities (*Count = 3, Support = 4*). Participants were also interested in the different types of modalities that drivers or passengers can choose to interact with IVIS (e.g., different modes of touch interaction, voice, or steering wheel control). In particular, they want to know which modality drivers primarily use (1) and whether this use differs across features (2,4).

INF-7: User Information (*Count = 2, Support = 3*). For user-specific information, respondents see value in comparing data from different regions (3, 4) or comparing data for different target groups (e.g., by demographics or frequently used features).

Regarding the visualizations proposed in Chapter 6, participants agreed that they already partially address the information needs **INF-1**, **INF-2**, **INF-3**, and **INF-5**. However, they do not provide system information (**INF-4**), information about different modalities (**INF-6**), or user information (**INF-7**).

7.2.4 Interaction Needs

To extract the interaction needs of the participants, we asked them to imagine a tool that would meet all their information needs and to explain how they would like to use this tool in their daily work. The expectations were very consistent, as they all expected to use the tool to define a new UI concept, to validate an existing and already implemented UI concept, and to question the customer value of a feature. Based on these insights, we then explored how users would like to interact with the anticipated tool and how they would like to configure it to meet their needs. The answers to these questions form the interaction needs. As shown below, we grouped the 14 individual needs into 4 categories.

INT-1: Task Definition (*Count = 4, Support = 10*). Participants emphasized that they want to configure their analytics based on individual use cases, rather than having a "one-size-fits-all" dashboard. While they valued certain standard metrics to be displayed, they wanted to define specific tasks or characteristics for which they needed detailed analytics. To define the tasks of interest, all participants (I1-I4) asked if it would be possible to interactively define sequences without having to manually enter the object identifiers. They suggested using a desktop-based version of IVIS, arguing that this would facilitate task definition since the UI software consists of thousands of elements. However, for known use cases, they suggested traditional input options such as drop-down menus to select UI elements as start and end points (I1, I2). Here, one participant (I2) mentioned that the analysis tool should use the same UI identifiers as those used in the UI concept description.

INT-2: Analysis (*Count = 5, Support = 13*). When it came to analyzing, participants were concerned about overall complexity, noting that traditional dashboards often tend to be overloaded and cluttered. Accordingly, they asked for features that would allow them to reduce the complexity of the results. They also wanted to be able to drill down through different levels of granularity depending on their use case, rather than being presented with all the results at once (I1, I2, I4). All participants argued that they need to be able to compare usage by system, context, and user information (I1-I4). Most of the proposed filtering options focused on system-specific information such as car type or software version.

INT-3: Operating Aids (*Count = 3, Support = 5*). Two participants (I1, I2) mentioned that the tool should be adaptable according to the user's expertise. They suggested that the tool could provide an *"exploration mode"* (I2) to help them explore the UI. They also asked for the possibility to display reduced versions of the plots proposed in Chapter 6.

INT-4: Sharing and Collaboration (*Count = 2, Support = 4*). Participants expressed the need to share the visualization with colleagues and decision makers, either in a portable format (I1-I3) or through a link that provides direct access (I3).

The visualizations presented in Chapter 6 are stand-alone visualizations without a user interface. Therefore, they do not address any of the identified interaction needs.

7.3 Introducing ICEBOAT

Study 1 identified the information and interaction needs of automotive UX experts for visualization and analysis of customer data and confirmed that the visualizations presented in Chapter 6 partially satisfy the information needs of UX experts. However,

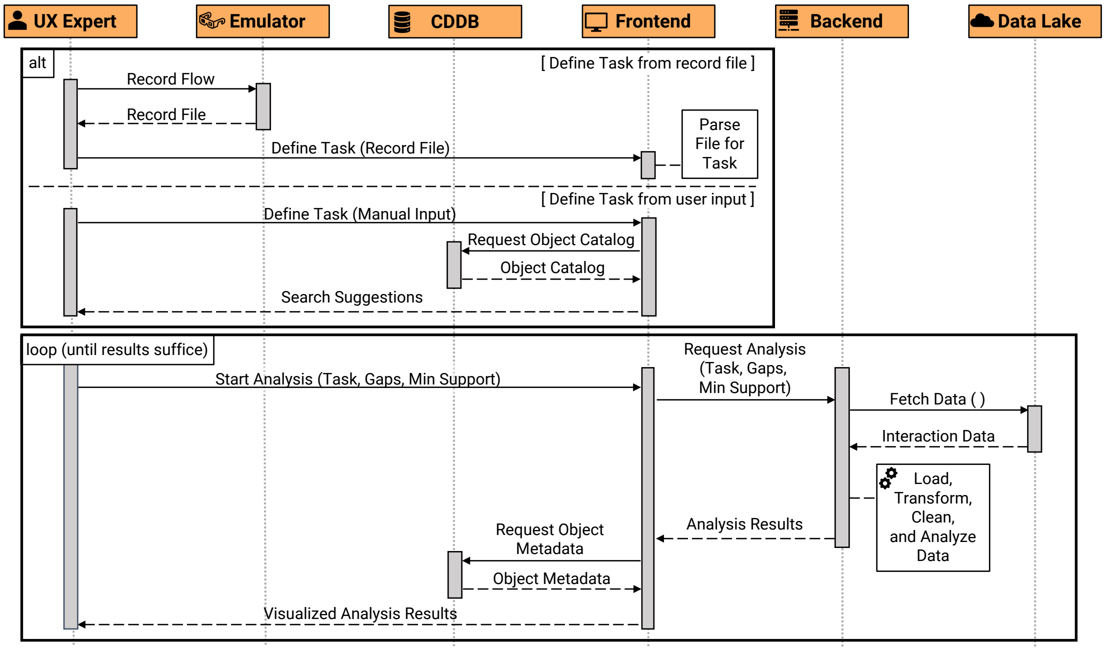


Figure 7.2: Sequence diagram depicting all processes and how they interact with each other.

they do not provide an interface that addresses the interaction needs. Therefore, following **INF-1** – **INF-7** and **INT-1** – **INT-4**, we developed ICEBOAT, an interactive user behavior analysis tool for automotive UIs. ICEBOAT refines the visualizations presented in Chapter 6, adds new functionalities and connects them in a meaningful way. Built on top of the Telematics Data Logging Framework and the User Behavior Evaluation Module introduced in Section 3.1, it automates task definition, data processing, and visualization generation, making large amounts of customer data easily accessible for UI evaluation.

We developed ICEBOAT using a co-design approach with four iterations. We invited the background interview participants as co-designers to each of the sessions, which were conducted remotely using Microsoft Teams.

7.3.1 System Architecture

ICEBOAT consists of a web-based frontend application for data visualization and a backend system for data processing (see Figure 7.2). The frontend, developed using the JavaScript framework *Vue.js*[1], receives data from three different services: The *Concept Database* (containing all UI information), the *IVIS Emulator* and the backend. The *IVIS Emulator* virtualizes the IVIS so that it can be executed on a computer as if it were running in the car.

The backend is divided into two services: An Application Programming Interface (API) service built with *FastApi*[2] web framework and a data service. The API service receives the analysis requests, passes them to the data service, and returns the results. The data

[1] https://vuejs.org
[2] https://fastapi.tiangolo.com/

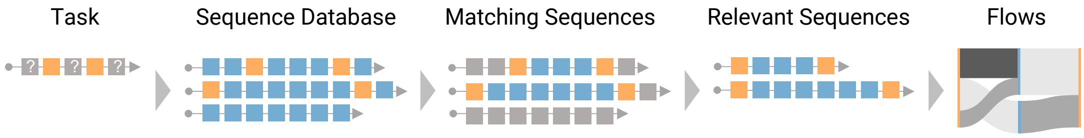


Figure 7.3: Sequence and flow extraction according to the task definition.

service uses *PySpark*[3] to efficiently extract, transform, and load the customer data stored in the data lake. The data lake is updated daily with the latest customer data. After running the analytical queries and extracting relevant user flows (see Figure 7.3), the backend returns the results to the API service. The frontend enhances the processed data with additional UI-specific information from the *Component Description Data Base (CDDB)*. We chose this architecture to make ICEBOAT easily extensible and to ensure interoperability (e.g., with another back end solution).

7.3.2 Interactive Web Application and IVIS Emulator

Figures 7.4 and 7.5 show an overview of the final tool after four iterations. The interface consists of a *Dashboard Page* and a *User Flow Analysis Page*. The user flow analysis page consists of four panels that fade in one after another based on the user's input addressing the need for a drill down mechanism to reduce complexity (**INT-1**). An overview of all components is given below:

1. **Dashboard Page:** Upon opening ICEBOAT, users are presented with a dashboard page (see Figure 7.4) that welcomes them and gets them started by explaining the purpose of the application. The tiles below the introduction report specific KPIs that describe the underlying data. For example, the number of trips on which the analysis is based and the number of logged interactions with the head unit. This page therefore onboards users and provides a perspective on the data to facilitate entry into the analysis (**INT-3**).
2. **Task Definition:** The tool provides two ways for users to define the task they want to analyze. (1) They can select the UI elements that define the start and end of a task from a searchable drop-down menu filled with all the UI elements that exist within the IVIS. (2) Alternatively, you can define a task using the *IVIS Emulator* (see Figure 7.4). ICEBOAT then automatically extracts all similar flows from the customer data and visualizes them. This provides the user with a playful and easy way to define tasks without having to know the naming conventions of specific UI elements. These two options address the interaction needs of **INT-1** and **INT-3**.
3. **Task Overview:** After loading the data for the specified task, the *Task Overview Panel* (see Figure 7.5) presents aggregations for all user flows between the start and end event of the task (**INF-3**). The data is presented as an adapted Sankey diagram (see Section 6.1.2) and in tabular form. The two views provide information about the average time between two consecutive interactions in a flow, the gestures used, the relative and absolute frequency of flows, the total number of interactions

[3] https://spark.apache.org/docs/latest/api/python/index.html

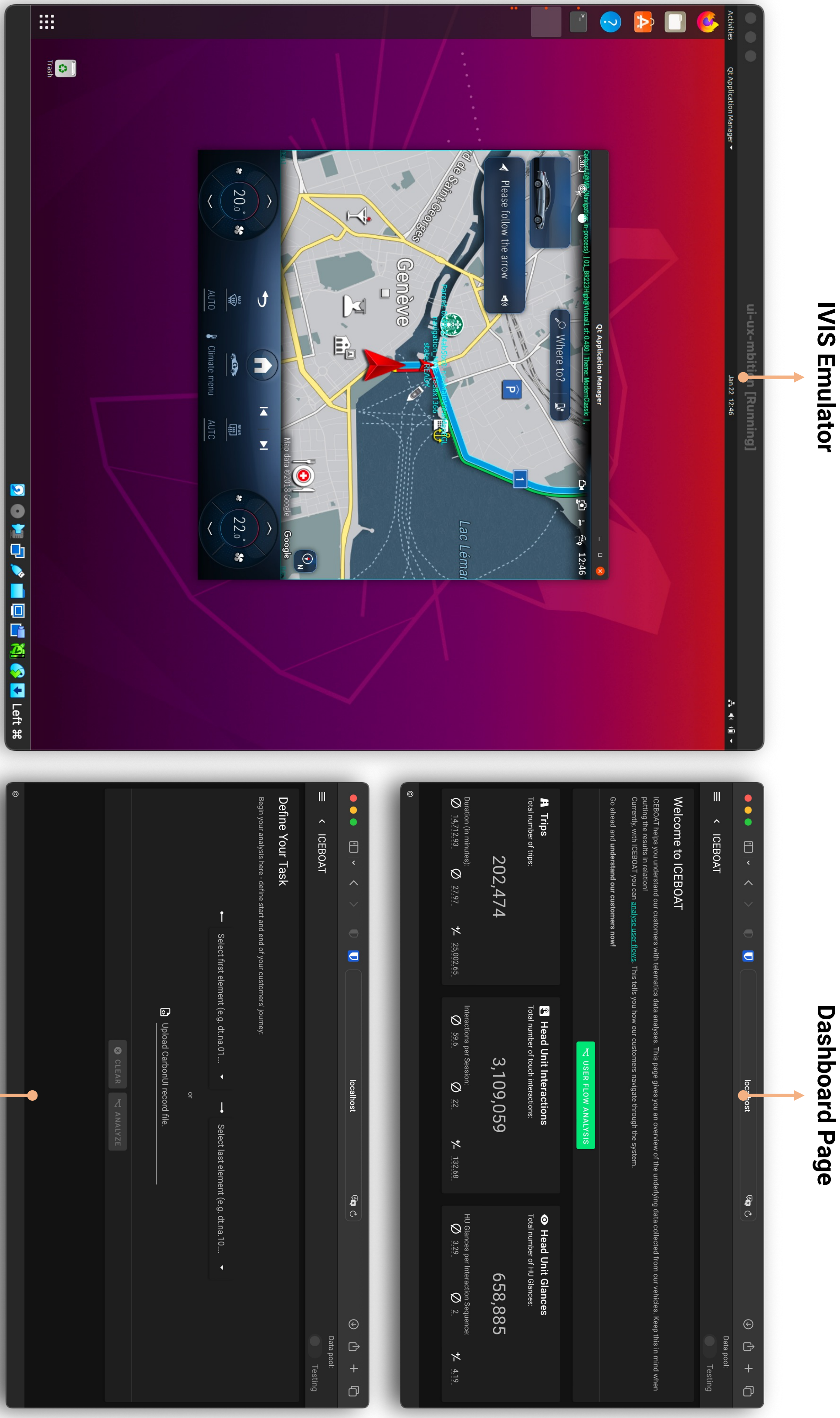


Figure 7.4: Overview of the IVIS Emulator (left), the Dashboard Page (top right) and the Task Definition Panel (bottom right).

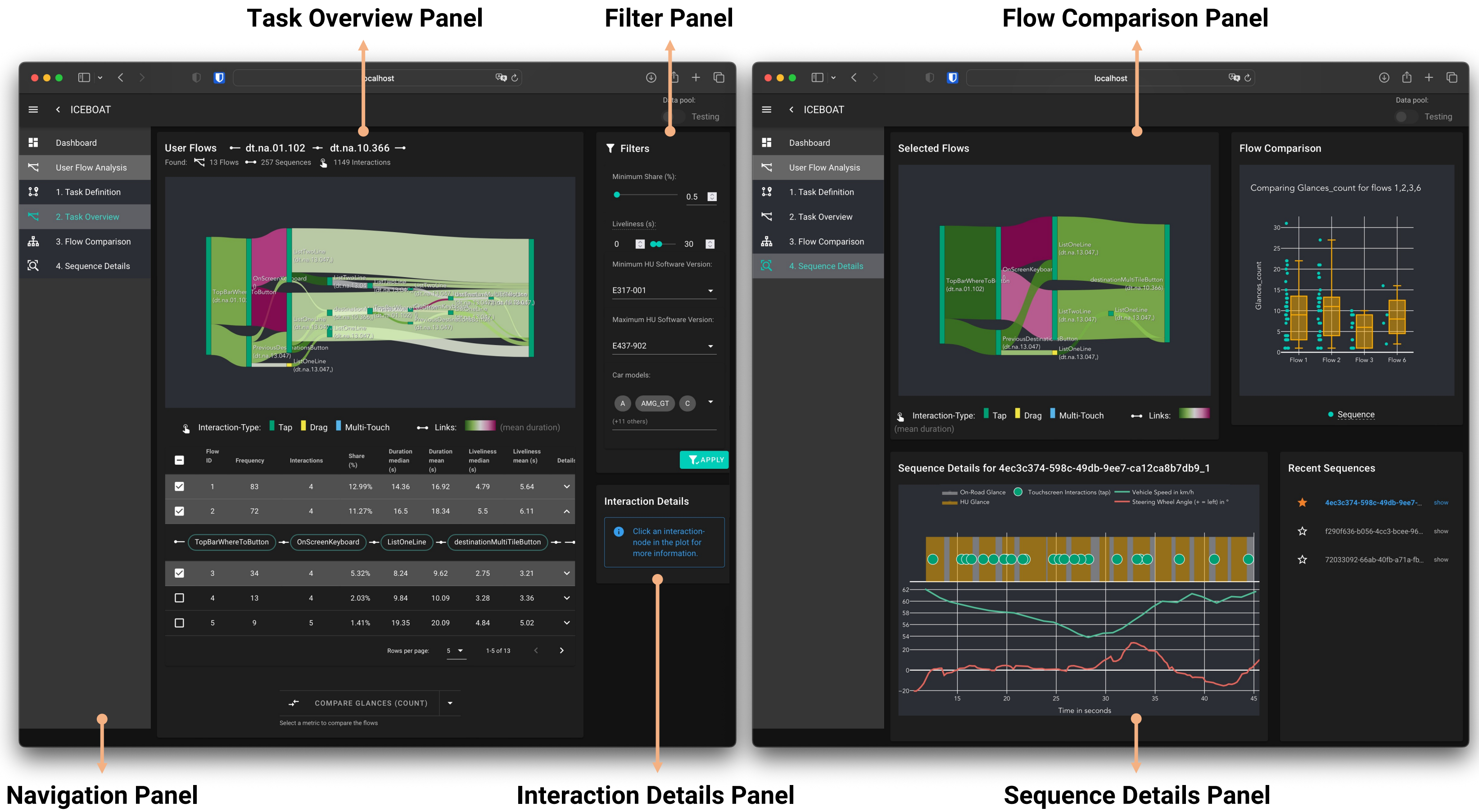


Figure 7.5: Overview of the User Flow Analysis Page The left part shows the initial Task Overview Panel. The Sankey diagram consists of all flows that satisfy the task and filter requirements. The table below the Sankey diagram shows various flow-related user behavior metrics. Clicking a UI element in the Sankey diagram opens the Interaction Details Panel, which displays the technical description of the element. On the right is the Flow Comparison panel, which consists of a reduced Sankey diagram and box plots showing the distribution of the total number of glances per interaction sequence. The Sequence Details Panel shows the touch interactions, gaze, and driving behavior of a specific sequence over time. The given example represents the use case of Study 2.

in a flow, and the average flow duration (**INF-1, INF-3**). The *Filter Panel* on the right-hand side allows users to customize the visualization to reduce visual clutter (**INT-2**). It also allows filtering for specific software versions or car types (**INF-4**). The visualized flows are based on all of the user interaction sequences collected from production vehicles that match the task definition (see Figure 7.3) and the applied filters.

4. **Flow Comparison:** The *Flow Comparison Panel* shows a reduced Sankey diagram of the selected flows (**INT-3**) and a box plot comparing each of the flows according to a selected metric, such as gaze duration or time on task (**INF-1**, **INF-3**). The box plots show the distribution of these values for all sequences contained in the flow. Thus, each dot represents a single sequence, which users can select (by clicking) to open the Sequence Details view for that interaction sequence.

5. **Sequence Details:** The *Sequence Details Panel* allows the user to explore details about a single interaction sequence. The view blends glance data (on-road, off-road, center stack touchscreen) with contextual driving data such as speed or steering angle, and embeds the touchscreen interactions (see Section 6.1.4 (**INF-5**). On the right, users see a history of sequences they have viewed and can save specific ones as favorites (**INT-3**).

7.3.3 How ICEBOAT Empowers Automotive UX Experts

The provided visualizations and analyses support UX experts in the design and evaluation of touchscreen-based IVIS. With ICEBOAT, we support practitioners in overcoming three key challenges related to the use of big data analytics in the design and evaluation of IVIS:

Data-Driven Decision-Making Due to cultural, organizational, and technological challenges, data plays only a minor role in the decision-making process related to automotive design [65]. As technology improves, large amounts of driving and interaction data are being collected. However, as pointed out in Chapter 4, automotive UX experts still report that they lack the tools to access and analyze the data directly and independently. Based on a telematics data processing framework, ICEBOAT gives UX experts permanent and immediate access to usage data collected live in the field. This allows UX experts to analyze interaction, glance and driving data independently throughout their workflow.

Automotive-Specific Analysis General-purpose tools for big data analytics are often disconnected from the workflows of domain experts [86]. This is also true in the automotive domain. UX experts report (Study 1) that current tools do not meet their specific needs for task definition, user flow exploration and comparison, and visualization of individual usage sequences. ICEBOAT supports the definition of user tasks, allows users to compare specific flows according to various usability and driver distraction related metrics, and enables UX experts to visualize details of individual interactions with the IVISs.

Information Overload UX experts are not trained to analyze large amounts of data. Therefore, data visualizations and interactions must be easy to understand and their benefits obvious to avoid information overload [83]. ICEBOAT allows UX experts to explore UI-relevant data without requiring technical knowledge. The IVIS emulator allows UX experts to easily define the scope of their analysis and pre-defined KPIs are visualized decoupled from the detailed user flow analysis to avoid information overload. The user flow analysis allows users to start with a broad overview and then zoom in on details as the analysis progresses. This drill-down concept (panels appear one after the other) fits the workflow of UX experts and presents only the information that is needed.

7.4 Study 2: Evaluation

To assess whether ICEBOAT enables UX experts to independently explore large-scale behavioral data, we conducted an evaluation study with UX researchers, designers, and data scientists.

7.4.1 Method

We conducted usability testing, interviews, and a Context of Use Questionnaire. We aimed for a representative sample of participants and used standardized measures to assess usability. To guide the evaluation, we followed a test plan that we created based on the recommendations of Still and Crane [244].

Participants We recruited 12 potential users from Mercedes-Benz and MBition, a Mercedes-Benz software hub: 4 designers, 4 UX researchers, and 4 data scientists. We included data scientists for two reasons: First, due to cross-functional development teams, data scientists often work closely with designers or UX researchers in decision-making processes. Second, because of their familiarity with data analysis, we expected data scientists to provide a different perspective and baseline for understanding data. We did not invite participants who had already participated in the study, as this could skew the results when evaluating usability [248, 249, 250]. The age of the participants ranged from 21 to 41 years (mean 29.6, SD 5.6) and their work experience from 0.5 to 20 years (mean 5.3, SD 5.8). All but one participant had a college degree.

Scenario and Evaluation Tasks To create a realistic evaluation environment, we derived a test scenario from the storyboard[4] resulting from Study 1. In this scenario, the UX experts are asked to evaluate the destination entry task of the navigation feature. They should use the IVIS Emulator to define a representative task and then analyze it for bottlenecks, driver distraction, and outliers in glance behavior.

Procedure We collected demographic data in a pre-survey before the experiment. At the beginning of the experiment, we introduced the scenario and asked the participants to complete seven evaluation tasks (compare Table 7.1) that resembled the scenario introduced above. First, we shared the IVIS emulator with the participants and asked them to complete the first task. Then we switched to the ICEBOAT screen. Before the next task, we had the participants practice thinking aloud by asking them to describe the dashboard page and give feedback. Users then navigated to the *User Flow Analysis* page of the tool and proceeded with the second task. During the test, participants were

[4]The storyboard is provided in the supplementary material: `http://kups.ub.uni-koeln.de/id/eprint/65348`

Table 7.1: Evaluation tasks.

No.	Task
1	Use the record button of the IVIS emulator to record a flow beginning at the navigation system's start screen and ending with the "Let's Go" button.
2	Use the record file to define and analyze the customer journey.
3	What are the top 5 flows (by share)? Use the filters to only display these flows.
4	Identify one bottleneck in the flows. Could you explain the potential causes?
5	Compare the glance behavior (count) of the first 3 flows.
6	Which sequence in flow 1 has the highest glance count?
7	Identify one long glance and explain the driving situation.

free to explore the tool and ask questions. After the participants completed all tasks, we collected their feedback both verbally and with the post-survey. We also recorded whether participants encountered any technical problems during the test. Since participants in a lab study usually answer usability questionnaires on-site [244], we had the participants fill out the surveys online immediately after the study. However, we stopped recording the interviews and turned off the cameras and microphones so that participants would not feel observed while completing the survey.

Measures We counted and coded the errors participants made while solving the tasks, and collected participants' feedback and interpretations of the visualization. We also had the participants fill out the System Usability Scale (SUS) questionnaire [48] and the Context of Use Questionnaire (see Table 7.2).

Table 7.2: The Context of Use Questionnaire.

No.	Question
1	I think the system allows me to analyze telematics data on my own (independent from another person or department)
2	I think the system makes telematics data accessible
The system provides insights into telematics data...	
3	...which are new to me
4	...that help me to better understand how our customers interact with the infotainment system
5	...which help me to observe how our customers interact with the infotainment system in different driving situations
6	...which allow me to base my decisions on data
7	...which help me resolve discussions about feature priorities
8	The system helps me to identify usability issues in our infotainment system
9	Having the system available would accelerate our current workflow with telematics data analysis
10	In which phase of the design process would you use the system?

Test Environment & Schedule Due to the distributed work environment, we conducted all experiments remotely using Zoom. With the users' permission, we recorded each test session to analyze the session afterwards and to quantify the error rates per task. We prepared a setup with the IVIS emulator open on one screen and the ICEBOAT tool open on another. This mimics the setup we imagine users would have when using the prototype in their day-to-day work. Using screen sharing, we allowed participants to remotely interact with the emulator and analysis tool, making our remote test environment as similar as possible to the production environment. We ran 12 tests of one hour each over three weeks.

7.4.2 Quantitative Results

We present results from the SUS and Context of Use Questionnaires, as well as additional qualitative insights.

SUS ICEBOAT received a mean SUS score of 68.125 (MD=70, SD=16.89), which, according to Lewis and Sauro [251], is average. While data scientists rated the tool with a mean score of 80, UX experts rated it with a mean score of 62. Cleland et al. [252] reports a similar spread between domain experts and data scientists. When evaluating the usability of the proposed big data analytics platform, data science testers rated the platform almost 20 points higher than policymakers (75.0 vs. 56.7).

Context of Use Questionnaire The mean score of the Context of Use Questionnaire was 4.2 out of 5 (MD=4.24, SD=0.33). In contrast to the SUS, only two questions were rated differently by data scientists and UX experts. The Pearson correlation between the results of the Context of Use and SUS questionnaires was not significant (R=0.55, p=0.061), suggesting that usability and value to the experts' workflow are not directly related.

7.4.3 Qualitative Feedback

Overall, participants found the tool valuable and easy to use. They reported that it would open up new possibilities for them and make their workflow much more efficient, *"I think this really makes our job easier, especially when you see how quickly you can get evaluations compared to how long it takes now"* (P3) **(INT-2)**. They also report that ICEBOAT provides effective insights because it *"[...] would provide better answers to many questions"* (P9). P9 further states: *"It is relatively difficult for us to make statements about groups that drive premium vehicles. With the tool you could get to those people."* **(INF-7)**. They also appreciated the ability to define the task using the IVIS Emulator, as it allows them to define the scope of their analysis without having to know the identifiers of specific UI elements **(INT-1)**. They report that this facilitates exploration and reduces the burden of using this tool. The design and layout of the tool was generally well received.

7.4.4 Data Understanding

ICEBOAT stimulated discussion about usability and safety improvements as participants solved the tasks. The Sankey diagram visualization was easy for participants to understand, and they were able to identify bottlenecks using the color scale or by manually comparing interaction times (shown when hovering over the flows) **(INF-3)**. One participant immediately suggested that the search suggestions could be improved to reduce the number of characters the driver has to type, because *"[t]he list keeps updating as you type. So it takes the user more time to find what they are looking for if they type more characters"* (P8). When comparing the top three flows based on the number of glances, 5 participants asked for clarification on how to interpret the box plots, but were able to identify the flow with the lowest average number of glances once explained **(INF-1)**. Participants quickly identified the flow with the most glances **(INF-1)** and appreciated the ability to select individual sequences to open the Sequence Details Panel **(INT-2)**. Using the Sequence Details Panel, they were able to assess the dependencies between glance, interaction, and driving behavior (**INF-5)**, *"The driver is on the move and slows down in the course of the interaction"* (P6), *"after brief glances at the road, the driver immediately performs several interactions"* (P10). However, participants interpreted the steering angle changes differently, with some interpreting them as a sign of distraction and others as a driving maneuver. Overall, participants found the tool helpful and argued that the insights can be particularly valuable in defining the scope of specific user studies to explore not only the *"what"* but also the *"why"*.

Errors In general, participants reported that they understood the tasks easily and were able to complete them efficiently. When interacting with ICEBOAT (Task 2-7), participants made only minor errors. For example, two participants initially chose a minimum support that was too high or too low, making the visualization either too cluttered or too sparse. Also, to create a reduced Sankey diagram in Task 5, four participants wanted to further reduce the flows using the minimum support instead of using the checkboxes in the table. Most of the errors occurred when interacting with the IVIS Emulator. While all participants successfully created a recording, only five out of twelve users did so with the expected start and end, as they did not start the recording on the expected screen. When asked, the participants stated that they thought they should start the recording directly from the main menu. However, this is more of a study-induced error with no practical implications. When asked to elaborate on their errors, participants stated that it takes some time to get used to the tool but *"[o]nce you get used to it and it's established as a working tool, it's super helpful"* (P12).

7.5 Limitations and Future Work

While our results show that ICEBOAT effectively empowers UX experts and meets most of the information and interaction needs for analyzing large amounts of usage data, some limitations should be considered. First, data scientists rated the usability of the tool higher than UX experts. This may be due to their experience with other data analysis tools. However, it also suggests that further research should be conducted to address the shortcomings with respect to users unfamiliar with data analysis. In addition, the slight delay and minor issues with the screen sharing and remote control feature may have influenced the results. Second, the study only considered touch interactions on the center stack screen. However, drivers can also interact with IVISs using speech or hardkeys. Thus, to satisfy **INF-6**, the next step would be to introduce these modalities in ICEBOAT. Furthermore, we only interviewed employees of one OEM. While our results as presented in Chapter 5 and related work [65] suggest that development practices and challenges are similar across most automotive OEMs, information and interaction needs may be skewed. Due to privacy concerns, we are not allowed to collect personal data. Thus, the only way to satisfy **INF-7** is to use a combination of available filters to define *"target groups"* (e.g., luxury car buyers vs. compact car buyers, as indicated by P9). Finally, we recorded the tests remotely, and 4 participants reported that the remote control function temporarily stopped working. While we were able to immediately restore control for 3 of the 4 people, this prevented one participant from completing a task. We had this participant verbally instruct us to complete the task and then restored remote control.

7.6 Conclusion

We present ICEBOAT, an interactive tool that makes millions of in-vehicle user interactions available to UX experts to effectively and efficiently visualize and evaluate drivers' touchscreen interactions with IVISs.

In Study 1, we identify the information and interaction needs of UX experts when analyzing large amounts of telematics data. Our findings reveal a conflict of interest: UX experts want to access as much data as possible and perform IVIS-specific analyses, but are deterred by the complexity of traditional big data visualization tools. ICEBOAT addresses this conflict of interest by (1) allowing users to define a task via a IVIS emulator, (2) automating all data processing and cleaning while still allowing manipulation of the metrics that matter, and (3) providing an interactive drill-down mechanism that allows users to start broad and zoom into the details of individual interactions. In Study 2, we show that UX experts and data scientists can effectively use ICEBOAT to visualize large amounts of automotive usage data to evaluate touchscreen-based IVISs. Most importantly, ICEBOAT empowers UX experts and contributes to the democratization of data in the automotive domain.

Part III

Modeling Driver Behavior

The studies presented in Part I show that UX experts are struggling to cope with the increasingly complex design space of IVISs. Part II presents visual analytics as a solution that can make improve the evaluation of IVISs. However, as outlined in Section 2.2.4, such visualization methods reach their limits when it comes to quantifying effects, exploring previously unknown and complex relationships, or making predictions about human behavior. As such, they are only a partial solution to the challenges UX professionals face when designing IVISs (see Part I). To design IVISs that are safe to use in all driving situations, we need to (1) understand how users adapt their behavior to different driving demands and (2) be able to evaluate prototypical designs for their distraction potential as early as possible. In this chapter, we address both of these issues. First, we present explanatory statistical models that quantify drivers' tactical and operational self-regulation and show how users adapt their behavior according to changes in driving demand. Second, we show how machine learning methods can be used to predict and explain the visual demand of in-vehicle touchscreen interactions, and how we envision using these predictions and explanations to be evaluate early-stage IVISs designs.

Chapter 8

Multitasking While Driving: How Drivers Self-Regulate Their Interaction with In-Vehicle Touchscreens in Automated Driving

Context To interact with touchscreen-based IVISs, drivers must divide their attention between the primary driving task and the non-driving-related secondary task on the center stack touchscreen. Although drivers are proven to self-regulate their secondary task engagements based on driving demands [25, 26, 27], this task-switching behavior is directly associated with an increased crash risk [126]. This is particularly critical as drivers tend to overestimate the capabilities of automated driving functions [28] potentially making it more likely to engage in non-driving-related tasks [24] in situations in which they are supposed to monitor these functions constantly [30]. Thus, a deep understanding of how drivers self-regulate their secondary task engagements in response to varying driving demands is inevitable to design IVISs that are safe to use in all situations. However, as of now, no research exists that investigates drivers' self-regulation on a level detail sufficient to draw conclusions about the UI design of touchscreen-based IVIS. However, the importance of being able to do so is reflected in the results presented in Chapter 4 and Chapter 5.

Contribution To better understand how drivers adapt their engagement in secondary touchscreen tasks, we investigate the effect of driving automation (manual vs. ACC vs. ACC + LCA), vehicle speed, and road curvature on drivers' tactical and operational self-regulation. We further show how the effect of driving automation depends on vehicle speed and road curvature. To evaluate tactical self-regulation, we fit generalized linear mixed models estimating the probability of drivers interacting with specific UI elements. Our results show that drivers self-regulate their interaction behavior differently across the UI elements. During ACC+LCA driving, the odds of a driver interacting with a map element are, for example, 1.62 times as high as for manual driving. The probability to interact with a regular button, however, remains similar. Furthermore, we measure drivers' operational self-regulation as glance behavior adaptions. The multilevel modeling results indicate that drivers adapt their glance behavior based on automation level, vehicle speed, and road curvature. Across all driving situations, the mean glance duration increases by 12 % for ACC driving compared to manual driving and by 36 % for ACC+LCA driving. The odds that drivers perform a glance longer than 2 seconds are 1.6 and 3.6 times as high, respectively.

Related Publications This chapter is adapted with minor changes from Ebel et al. [4, 7].

8.1 Study Design

We identify two main research gaps in the current state of the art: (1) Current work is mainly focused on self-regulation when interacting with mobile phones or when engaging in general secondary tasks such as eating, drinking, or talking to a passenger. No work addresses operational and tactical self-regulatory behavior during explicit interactions with IVISs. (2) Whereas multiple studies investigate the general effect of partial automation on drivers' self-regulation, there is yet no detailed investigation on the interdependencies between driving automation, vehicle speed, and road curvature.

Considering that modern IVISs are increasingly complex and incorporate nearly all the functionality of smartphones and that ACC and LCA are becoming more capable and accessible, we argue that both aspects need to be examined in more detail. Therefore, we aim to answer the following research questions:

RQ1: To what extent do drivers self-regulate their behavior on the tactical level when engaging in secondary touchscreen tasks depending on driving automation, vehicle speed, and road curvature?

RQ2: To what extent do drivers self-regulate their behavior on the operational level when engaging in secondary touchscreen tasks depending on driving automation, vehicle speed, and road curvature?

RQ3: Does the effect of driving automation on drivers' operational self-regulation vary in response to different driving situations?

8.1.1 Dataset

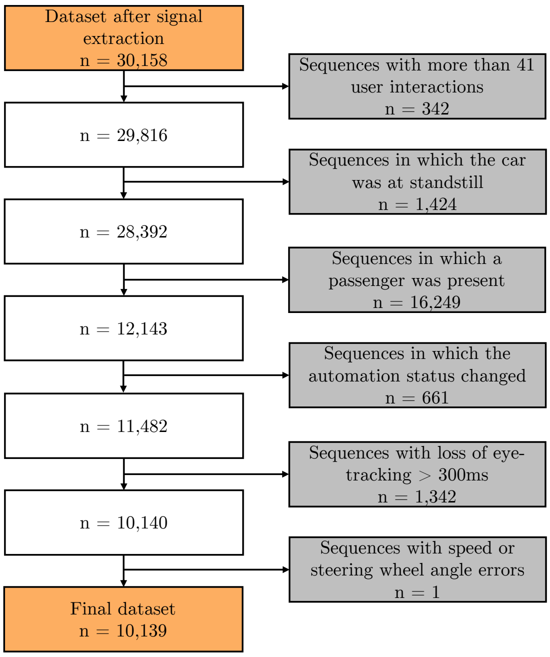


Figure 8.1: Data filtering procedure where n indicates the number of secondary task engagements.

More than 100 test vehicles contributed to the data collection from mid-October 2021 to mid-October 2022. The data was collected using the Telematics Data Logging Framework introduced in Section 3.1. After the data was logged, anonymized and stored, each signal was further processed according to the data processing and sequence extraction procedure described in Section 3.2.

After individual signal extraction, the dataset contained 98,038 sequences. To improve data quality and control of confounding factors, we applied strict exclusion criteria as visualized in Figure 8.1. We discarded all sequences with more than 41 interactions, which corresponds to the 99^{th} percentile of the distribution of interactions per sequence. We also discarded all sequences in which the car was at standstill. We filtered these sequences, because we are only interested in self-regulation while driving. To control for potential distractions or interactions by the front passenger, we deleted all sequences in which the front passenger seat belt buckle was latched. We also discarded all sequences in which the automation level could not be clearly assigned. This includes sequences in which the driver selected another automation level or overwrote the current level by accelerating or braking. The driving automation can also be deactivated due to external factors like a loss of lane marking or bad weather conditions. Furthermore, all sequences with a loss of eye tracking larger than $300\,ms$ were deleted. We also discarded all sequences during which the driver did not perform a gaze transition between the center stack touchscreen and the road. As we are interested in drivers' self-regulative behavior, we only consider sequences during which such regulation happened. Lastly, all sequences with errors in the speed or steering wheel angle signal were discarded. The final dataset[1] contains 31,378 sequences of which 18,449 are manual driving, 1,542 are ACC driving, and 11,378 are ACC+LCA driving.

8.1.2 Statistical Modeling

As stated in Section 8.1, we investigate how drivers' tactical and operational self-regulation changes in response to different levels of driving automation and driving contexts. In the following, we introduce the dependent and independent variables, and the statistical models we use. We define statistical significance at the level of $\alpha = 0.05$.

8.1.2.1 Dependent Variables

We chose following dependent variables to model tactical and operational self-regulation:

UI Interactions Current approaches are mostly investigating tactical self-regulation by comparing the likelihood of a driver engaging in a specific secondary task given different driving situations. We aim to investigate drivers' tactical self-regulation in greater detail, such that we can draw conclusions about the UI design itself. Therefore, we choose the *number of interactions* (discrete), the *number of touch gestures* (discrete) and the *probability of driver interactions with specific UI elements* (categorical) as dependent variables. The different categories of UI elements and touch gestures were introduced are given in Table 3.1 in Section 3.2.1. We did not consider interactions with the browser as they are blocked during driving.

[1] More detailed statistics are given here: http://kups.ub.uni-koeln.de/id/eprint/65348

Mean Glance Duration The *mean glance duration* is a continuous variable. It is computed as the sum of the duration of all glances toward the center stack touchscreen during a sequence divided by the total number of glances per sequence.

Long Glance The dichotomous variable *long glance* indicates whether a driver glanced at the center stack touchscreen for more than two seconds. Eyes-off-road glances longer than two seconds are associated with an increased crash risk [16]. The proportion of such long glances is an important factor in evaluating drivers' operational self-regulation.

8.1.2.2 Independent Variables

The dependent variables are analyzed with respect to the following independent variables:

Automation Level The automation level is a categorical variable with three distinct levels: *manual*, *ACC*, and *ACC+LCA*. According to SAE J3016 [30], these levels correspond to Level 0, 1, and 2 of driving automation. The automation level is constant throughout each sequence.

Vehicle Speed The vehicle speed is a categorical variable with three levels: $0\,km/h < v \leq 50\,km/h, 50\,km/h < v \leq 100\,km/h\,, v > 100\,km/h$. It is computed as the mean speed across a sequence.

Road Curvature The road curvature is a categorical variable with two levels: *straight* or *curved*. An interaction sequence is classified as curved if the maximum absolute steering wheel angle is greater than 50°or if the absolute mean steering wheel angle is greater than 5°.

8.1.2.3 Models

To account for the hierarchical data structure and the unbalanced study design we use mixed-effects models. Our data structure is hierarchical because interaction sequences are nested within trips and many trips occur within specific car types. Furthermore, not all combinations of the independent variables are observed in all trips and car types. This results in an unbalanced study design. However, mixed-effects models also referred to as multilevel models [108], are well suited for unbalanced designs and account for grouping hierarchies [253]. Thus they are well suited to test our hypotheses.

We performed all our analyses using R Statistical Software (v4.2.1) [254]. We used the *lme4* package (v.1.1.31) [255] to build the multilevel models, obtained p-values via the *lmertest* package (v.3.1.3) [256], and computed the pairwise post-hoc tests using the *emmeans* package (v.1.8.2) [257]. Regression tables were generated using the *stargazer* package (v.5.2.3) [258].

User Interaction Models To assess tactical self-regulation, we model the driver's decision to engage in a particular task in a particular driving situation. Specifically, we model the probability of drivers interacting with a particular UI element and the number of interactions and gestures drivers perform when interacting with the center stack touchscreen. To estimate the probability of a driver to engage with one of the UI elements, we fit one logistic mixed-effects model with random intercepts for each type of UI element and type of gestures. None of the two-way or three-way interactions were significant or proved to significantly improve the predictive performance compared to the additive model. We therefore omit these interaction effects.

To model the number of interactions and gestures that drivers perform during an interaction sequence, we fit two negative binomial mixed-effects models with random intercepts. We use negative binomial models because the number of interactions is a discrete count value. We could have also used Poisson models but our tests have shown that they suffered from overdispersion.

For all user interaction models we include *automation level*, *vehicle speed*, and *road curvature* as fixed effects. Furthermore, we include the trip during which the sequence was recorded and the car type as random effect.

Glance Behavior Models To estimate the *mean glance duration*, we fit four linear mixed-effects models with random intercepts. An exploratory data analysis showed that the distribution of the mean glance duration is heavily right-skewed. To satisfy the model assumption of normally distributed residuals we, therefore, apply a log transformation. In Model 1 we estimate the effect of driving automation on the mean glance duration across all driving situations by only selecting the automation level as a fixed effect. To account for the hierarchical structure of our data we include the trip during which an interaction sequence was recorded and the car type as random effects for both models. In Model 2 we add the vehicle speed and road curvature as additional fixed effects and allow for interaction effects. Similar to Model 1, the trip and car type are included as random effects. To estimate drivers' long glance probability, we fit two logistic mixed effect models with random intercepts. In Model 3 we select the automation level as a fixed effect and in Model 4 we add the vehicle speed and road curvature as fixed effects and model all interactions between the independent variables. The trip and car type information are, again, entered as random effects.

Visual inspection of residual plots and Q-Q plots of the final models did not reveal any obvious deviations from homoscedasticity or normality. We use Satterthwaite's degrees of freedom approximation to obtain p-values and evaluate significances [259]. For the post-hoc pairwise comparisons we use Tukey's multiple comparison method [260].

8.2 Results

In the following, we present the results obtained by fitting the above-introduced models to the 31,378 interaction sequences. By doing so we can model tactical and operational self-regulation. The analysis of the model coefficients and post-hoc tests allow us to quantify how drivers adapt their multitasking behavior according to changes in speed, road curvature, and driving automation.

8.2.1 Tactical Self-Regulation

In the following, we present the user interaction models and assess how users self-regulate their interactions with the center stack touchscreen on the tactical level.

Number of Touch Interactions and Touch Gestures Table 8.1 shows the parameters of the user interaction models. We modeled the number of touch interactions, tap, drag, and multitouch gestures per sequence. The results suggest that driving automation, vehicle speed, and road curvature affect the number of touchscreen interactions and gestures that drivers perform when engaging with the center stack touchscreen. The influence of the independent variables is generally similar but differs significantly in magnitude comparing *Tap* gestures to *Drag* and *Multitouch* gestures.

The β coefficients of the negative binomial model are given on a logarithmic scale. They can be interpreted as follows: Keeping everything else constant, an increase of one level in the predictor variable results in a e^{β} increase of the dependent variable. Thus, drivers perform $e^{0.11} \approx 1.12$ as many interactions during ACC driving and $e^{0.16} \approx 1.17$ as many interactions during ACC+LCA driving compared to manual driving. This corresponds to an increase of 12 % and 17 % respectively. Considering the different gestures that add up to the number of interactions, the modeling results suggest that, during automated driving, drivers in particular perform more drag or touch gestures compared to regular tap gestures. For example, during ACC+LCA driving the number of *Tap* gestures per sequence increases by 7 % whereas the number of *Drag* and *Multitouch* gestures increases by 73 % and 60 % respectively.

Road curvature also significantly affects the number of interactions and gestures that drivers perform on the center stack touchscreen. During curved driving, drivers perform $e^{-0.17} \approx 0.84$ as many interactions compared to straight driving. Wheres they only perform 12 % less *Tap* gestures, the number of *Drag* and *Multitouch* gestures reduces by 34 % and 28 % respectively.

The effect of the vehicle speed on the number of interactions and gestures is in general smaller compared to the effect of driving automation and road curvature. The results indicate that drivers do not, or only slightly, adapt their tap and multitouch behavior in response to changes in vehicle speed. However, the number of *Drag* gestures that drivers perform is significantly higher when driving at speeds above 50 km/h compared to driving at speeds of 50 km/h and below.

Type of UI Elements Table 8.2 shows the parameters of the user interaction models for all UI elements that occur in more than 10 % of all sequences[2]. The models were fit to predict the probability that a driver interacts with a specific UI element given the automation level, vehicle speed and road curvature. The results suggest that drivers adapt their interaction behavior with the center stack touchscreen based on automation status, vehicle speed, and road curvature. However, these effects do significantly differ for different types of UI elements.

The β coefficients for the independent variables given in Table 8.2 represent log-odds ratios. This means that, keeping everything else constant, a change in the predictor by one level results in a e^{β} increase or decrease in the odds that the driver interacts with the respective UI element. Considering the *Map* model the coefficients can be interpreted as follows: During ACC+LCA driving the odds that a driver performs a map interaction are $e^{0.48} \approx 1.62$ times as high as the odds of performing the same interaction in manual driving. On the other hand, when driving in curved conditions, the odds that the driver

[2]The results of the other models are provided here: `http://kups.ub.uni-koeln.de/id/eprint/65348`

Table 8.1: Negative binomial mixed-effects models that describe the number of touchscreen interactions, tap gestures, drag gestures, and multitouch gestures during an interaction sequence. For each model, the intercept and the coefficients describe the effect of the independent variables. They are shown along with the estimated standard error. The coefficients and standard errors of the negative binomial mixed-effects model are given on a logarithmic scale.

	Dependent variable:			
	Num. Interactions	Num. Tap Gestures	Num. Drag Gestures	Num. Multitouch Gestures
Intercept	1.74*** (0.01)	1.52*** (0.02)	−1.12*** (0.03)	−2.77*** (0.09)
Automation Level				
Manual†				
ACC	0.11*** (0.02)	0.07** (0.03)	0.29*** (0.06)	0.28** (0.09)
ACC+LCA	0.16*** (0.01)	0.07*** (0.01)	0.55*** (0.03)	0.47*** (0.05)
Vehicle Speed				
0-50†				
50-100	0.04*** (0.01)	0.03* (0.01)	0.12*** (0.03)	0.03 (0.04)
100+	0.01 (0.01)	−0.01 (0.01)	0.15*** (0.04)	0.04 (0.05)
Road Curvature				
straight†				
curved	−0.17*** (0.01)	−0.13*** (0.01)	−0.42*** (0.04)	−0.33*** (0.05)
Akaike Inf. Crit.	173,027.10	163,962.60	69,556.59	53,842.50
Bayesian Inf. Crit.	173,102.30	164,037.80	69,631.77	53,917.68

Note: † indicates the reference group , *p<0.05; **p<0.01; ***p<0.001

Table 8.2: Generalized linear mixed-effects models describing the probability of the driver interacting with Tab, List, Button, Homebar, or AppIcon UI elements during an interaction sequence. For each model, the intercept and the coefficients describe the effect of the independent variables. They are shown along with the estimated standard error. Coefficients and standard errors correspond to log odds ratios.

	Dependent variable:					
	Tab	List	Map	Button	Homebar	AppIcon
Intercept	−1.60*** (0.05)	−0.98*** (0.05)	−2.40*** (0.09)	−0.52*** (0.04)	−0.04 (0.05)	−1.37*** (0.05)
Automation Level						
Manual†						
ACC	0.12 (0.08)	0.06 (0.07)	0.18 (0.10)	0.05 (0.06)	−0.01 (0.07)	0.04 (0.08)
ACC+LCA	−0.07 (0.04)	0.25*** (0.04)	0.48*** (0.05)	0.15*** (0.03)	−0.17*** (0.04)	−4 (0.04)
Vehicle Speed						
0-50†						
50-100	−0.01 (0.04)	0.08* (0.03)	0.07 (0.05)	0.02 (0.03)	−0.01 (0.03)	0.06 (0.04)
100+	−0.18*** (0.05)	0.09* (0.04)	0.13* (0.06)	−0.04 (0.04)	−0.08 (0.04)	0.02 (0.04)
Road Curvature						
straight†						
curved	−0.10* (0.04)	−0.13*** (0.04)	−0.26*** (0.05)	−0.02 (0.03)	−0.03 (0.04)	−0.09* (0.04)
Akaike Inf. Crit.	29,284.14	38,230.86	29,078.23	41,130.21	41,482.00	32,765.17
Bayesian Inf. Crit.	29,350.98	38,297.69	29,145.06	41,197.04	41,548.83	32,832.00

Note: † indicates the reference group , *p<0.05; **p<0.01; ***p<0.001

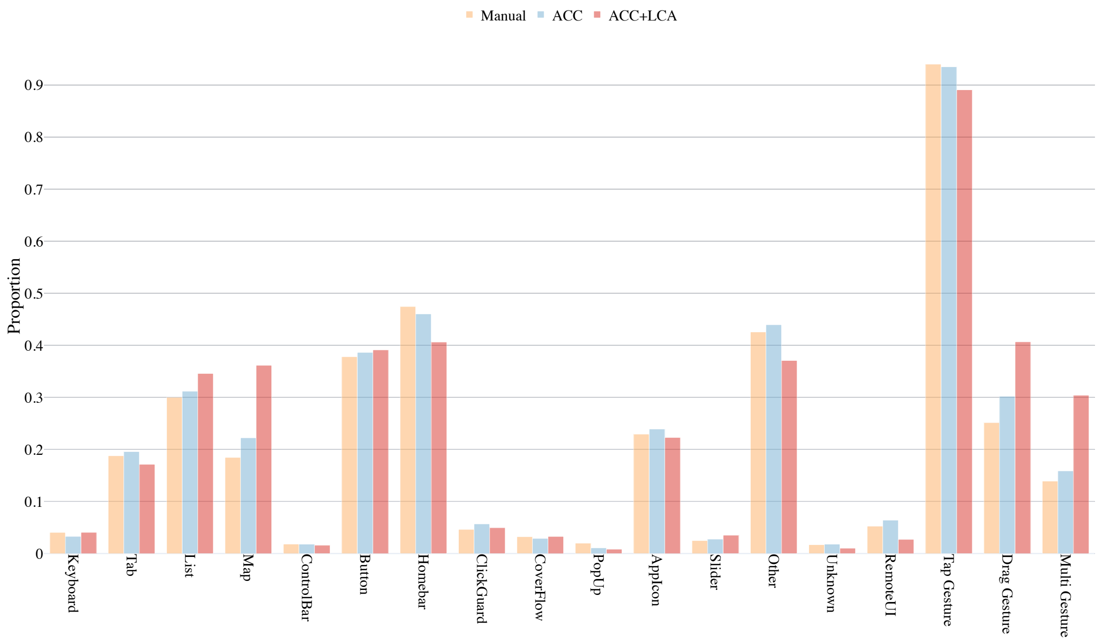


Figure 8.2: Bar plot showing the proportion of sequences in which the drivers interacted with a respective UI element and performed a specific gesture.

interacts with the map are $e^{-0.26} \approx 0.77$ the odds of performing a map interaction in straight driving conditions.

Whereas the effect of ACC is not significant for any of the models, the effect of ACC+LCA is significant for all Models except of the *Tab* and *AppIcon* models. While drivers are more likely to interact with *List*, *Map*, and *Button* elements, they are less likely to interact with the *Homebar*. The odds to interact with the homebar are $e^{-0.17} \approx 0.84$ the odds compared to manual driving. These effects are also shown in Figure 8.2.

Concerning the effect of vehicle speed, the effect of 50–100 is only significant for *List* interactions, suggesting that drivers perform more list interactions when driving between 50 km/h and 100 km/h compared to driving at speeds equal to or below 50 km/h. The effect of 100+ is, however significant for *Tab*, *List*, and *Map*. Whereas the odds of drivers interacting with Tab elements are $e^{-0.18} \approx 0.84$ times the odds of performing the same interactions at speeds between 0 km/h and 50 km/h. In contrast, for *List* and *Map* interactions the odds are 1.09 and 1.14 times higher.

The effect of road curvature is significant in all models but the *Homebar* and *Button* models. The coefficients suggest that during curved driving, drivers are in general less likely to interact with the center stack touchscreen. The odds for a driver to interact with these elements in curved driving conditions are between 0.77 and 0.91 the odds compared to straight driving.

Across all models, our results suggest that the effect of ACC+LCA driving on tactical self-regulation is larger than the effect of vehicle speed or road curvature. Whereas the tendencies for ACC driving are similar, the effect proves to be not significant ($p > 0.05$ for all models). Furthermore the effect of ACC+LCA driving is largest for list and map interactions and small or even negative for the other UI elements.

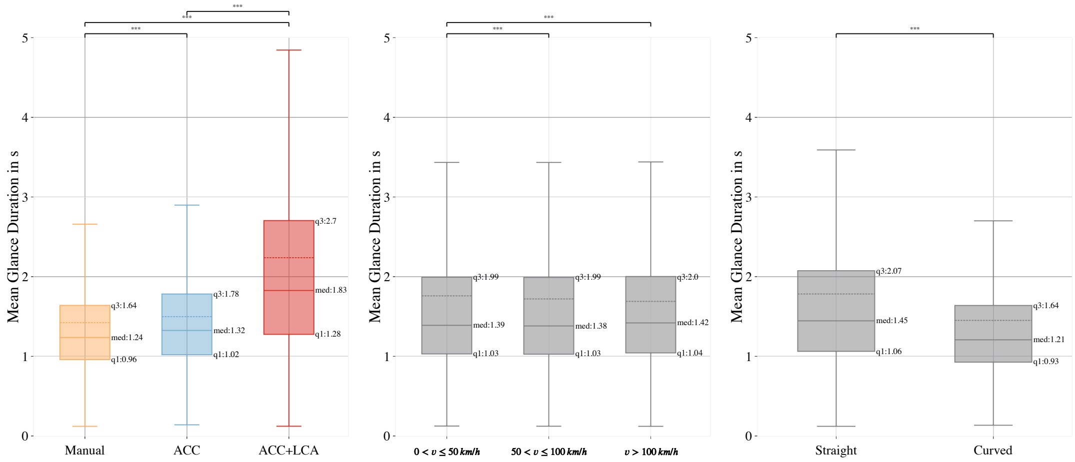
Table 8.3: Mixed-effects models for mean glance duration and long glance probability toward the center stack touchscreen. The coefficients and standard errors of the mean glance duration models are given on a logarithmic scale. The coefficients and standard errors for the long glance model represent log odds. All coefficients are shown along with the estimated standard error.

	Dependent variable:			
	Mean Glance Duration		Long Glance	
	linear mixed-effects		*generalized linear mixed-effects*	
	Model 1	Model 2	Model 3	Model 4
Constant	7.15*** (0.01)	7.25*** (0.01)	−0.25*** (0.05)	0.21*** (0.06)
ACC	0.10*** (0.01)	0.03 (0.04)	0.44*** (0.07)	0.11 (0.19)
ACC+LCA	0.31*** (0.01)	0.39*** (0.02)	1.29*** (0.04)	1.29*** (0.09)
50-100		−0.11*** (0.01)		−0.48*** (0.05)
100+		−0.17*** (0.01)		−0.66*** (0.06)
curved		−0.09*** (0.01)		−0.55*** (0.06)
ACC:50-100		0.12** (0.04)		0.39 (0.22)
ACC+LCA:50-100		−0.04* (0.02)		0.14 (0.10)
ACC:100+		0.15*** (0.04)		0.68** (0.21)
ACC+LCA:100+		−0.08*** (0.02)		0.17 (0.11)
ACC:curved		−0.03 (0.06)		0.34 (0.33)
ACC+LCA:curved		−0.15*** (0.04)		−0.54** (0.19)
50-100:curved		−0.03 (0.02)		−0.02 (0.09)
100+:curved		0.01 (0.03)		−0.35* (0.18)
ACC:50-100:curved		−0.07 (0.08)		−0.96* (0.43)
ACC+LCA:50-100:curved		0.07 (0.04)		0.05 (0.23)
ACC:100+:curved		−0.11 (0.10)		−0.74 (0.56)
ACC+LCA:100+:curved		0.10 (0.06)		0.87** (0.30)
Akaike Inf. Crit.	42,903.57	42,246.95	38,850.68	38,392.18
Bayesian Inf. Crit.	42,953.69	42,422.38	38,892.45	38,559.26
Note:				*p<0.05; **p<0.01; ***p<0.001

8.2.2 Operational Self-Regulation

Operational self-regulation is evaluated by identifying how drivers adapt their glance behavior. We measure glance behavior in terms of mean glance duration and long glance probability. The results of our (generalized) linear mixed-effects models (see Table 8.3) suggest that drivers adapt their glance behavior while interacting with the center stack touchscreen based on automation status, vehicle speed, and road curvature.

Mean Glance Duration The results of Model 1 as shown in Table 8.3 suggest that the effect of ACC and ACC+LCA on drivers' mean glance duration toward the center stack touchscreen is significant ($p < 0.001$) compared to manual driving. As the mean glance duration is measured on a logarithmic scale, the exponent of models' coefficients can be interpreted roughly as percent changes. When ACC is active, drivers' mean glance duration increases by $e^{0.10} \approx 1.11 = 11\,\%$. When ACC and LCA are both active, drivers' mean glance duration increases by 36 % compared to manual driving. Post-hoc testing using Tukey's pairwise post-hoc tests reveals that the difference between ACC and



(a) Driving Automation (b) Vehicle Speed (c) Road Curvature

Figure 8.3: Boxplots of the mean glance duration toward the center stack touchscreen representing the main effect of driving automation, vehicle speed, and road curvature. Statistically significant differences according to Tukey's pairwise post-hoc test are indicated as: *p<0.05; **p<0.01; ***p<0.001.

ACC+LCA is also significant. The effects are shown in Figure 8.3a. Figure 8.3 also shows the mean glance duration for different speed ranges (Figure 8.3b) and road curvature (Figure 8.3c). According to the modeling results [3], drivers' mean glance duration decreases by 6 % when driving between 50 km/h and 100 km/h and by 8 % when driving faster than 100 km/h compared to driving between 0 km/h and 50 km/h. It needs to be noted, that whereas these differences are statistically significant ($p < 0.001$) they are not observable in Figure 8.3b. This is because most of the correlation in the data is explained by the combination of fixed and random effects (trip and car type) rather than by the fixed effect (vehicle speed) alone. This means that the effect of the vehicle speed is only significant when taking into account trip and car type information. However, Figure 8.3b only shows the mean glance duration according to the vehicle speed. Our results further show that most of the variance in the data is explained by variations in the trip identifier. Considering that vehicle speeds of 0-50 km/h occur in urban driving but also in very controlled scenarios in a traffic jam on the highway, the trip identifier might be a proxy for different kinds of trips. This also shows that vehicle speed alone might not be the best indicator for changes in driving demand.

In addition to Model 1, Model 2 adds vehicle speed, road curvature, and the accompanying interactions as fixed effects. In this model, the combination of manual and straight driving, at speeds between $0 - 50\, km/h$ serves as a reference and all coefficients displayed in Table 8.3 need to be interpreted accordingly. Apart from the significant main effects for ACC+LCA, 50–100, 100+, and curved, the interactions between both levels of driving automation and vehicle speed and the interaction between ACC+LCA and curved are significant. Whereas the interaction effects of ACC and vehicle speed while driving straight are positive, they are slightly negative for ACC+LCA and vehicle speed. This means that the effect of ACC+LCA decreases slightly for higher speeds during straight

[3]Compare Model 5 in the supplementary material: http://kups.ub.uni-koeln.de/id/eprint/65348

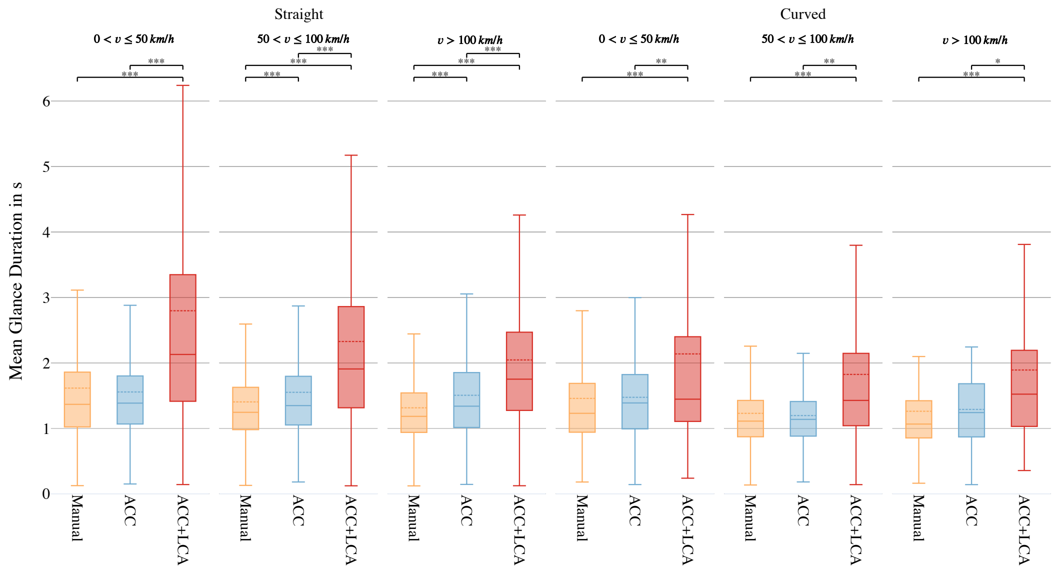


Figure 8.4: Boxplots of the mean glance duration toward the center stack touchscreen grouped according to road curvature (left and right half), vehicle speed (combination of three boxplots each), and driving automation (by color). Statistically significant differences according to Tukey's pairwise post-hoc test are indicated as: *p<0.05; **p<0.01; ***p<0.001.

driving whereas the effect of ACC increases with the speed for straight sequences. This can also be observed in Figure 8.4.

Furthermore, we are interested in whether the effect of ACC and ACC+LCA driving on drivers' self-regulation differs depending on the driving situations. We, therefore, perform pairwise post-hoc comparisons as shown in Figure 8.4. We adjust p-values based on Tukey's method for comparing a family of three estimates.

Drivers' mean glance duration is significantly higher during ACC+LCA driving compared to manual driving and ACC driving across all driving situations. During straight driving the mean glance duration during ACC+LCA driving compared to manual driving increases by 47 % (0-50 km/h), 42 % (50–100 km/h), and 36 % (100+ km/h). A similar but slightly smaller effect can be observed during curved driving. Here the mean glance duration increases by 27 % (0-50 km/h), 30 % (50–100 km/h), and 29 % (100+ km/h).

The effect of ACC driving compared to manual driving is only significant for straight driving sequences at speeds between 50 km/h to 100 km/h and at speeds above 100 km/h. For these two conditions drivers' mean glance duration increases by 15 % and 19 % respectively. During curved driving no significant effect can be observed for ACC driving.

Long Glance Probability The results of Model 3, as presented in Table 8.3, suggest that the level of driving automation significantly affects the probability that a driver performs a long glance during an interaction sequence. Both, ACC and ACC+LCA lead to an increase in the long glance probability.

The odds that a driver performs a long glance toward the center stack touchscreen are $e^{0.44} \approx 1.6$ (ACC) and $e^{1.29} \approx 3.6$ (ACC+LCA) times higher compared to manual driving.

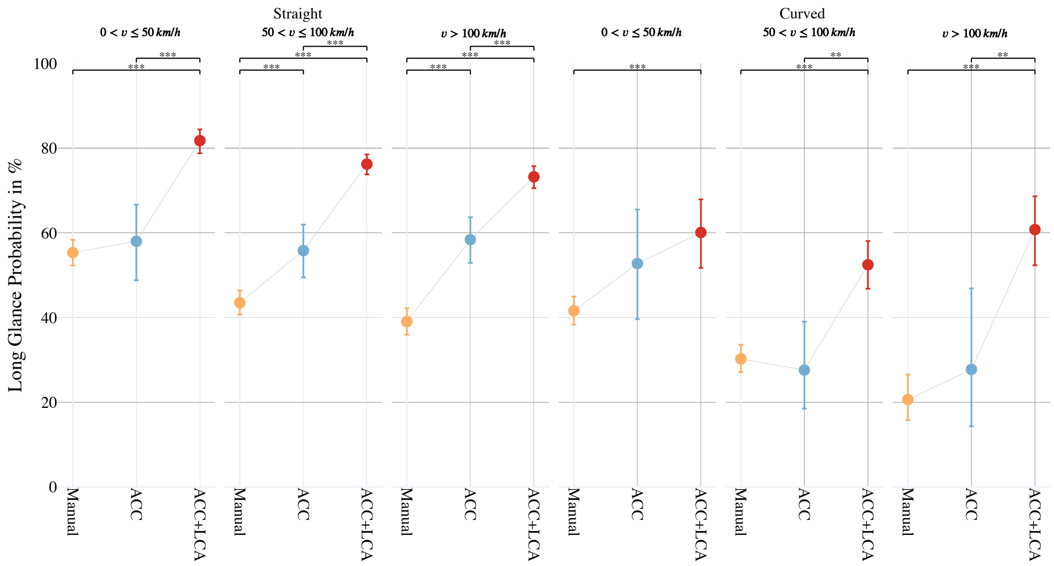


Figure 8.5: Marginplot of the predicted long glance probabilities and accompanying confidence intervals. The plots are grouped according to road curvature (left and right half), vehicle speed (combination of three boxplots each), and driving automation (by color). Significant results according to Tukey's pairwise post-hoc test are indicated as: *p<0.05; **p<0.01; ***p<0.001.

Post-hoc pairwise comparisons also reveal a significant difference between ACC and ACC+LCA with the odds being 2.4 times higher ($p < 1$) in the ACC+LCA condition.

The results of Model 4 show significant effects of vehicle speed, road curvature, and various interactions. Comparing the main effects, we observe that compared to the reference, the effect of ACC+LCA is roughly twice as high as the effects of 50–100, 100+, or curved. Furthermore, the effect of ACC driving alone is not significant but various of its interaction effects are. The model predictions and confidence intervals are visualized in Figure 8.5. Post-hoc tests comparing the different levels of driving automation for the different combinations of speed and road curvatures were performed using Tukey's multiple comparison method.

For interactions during straight driving at 0-50 km/h we observe a significant increase ($p < 0.001$) in the long glance probability during ACC+LCA driving compared to manual driving and ACC driving. The difference between manual driving and ACC driving is not significant. However, when driving at speeds between 50–100 km/h and speeds above 100 km/h, the long glance probability is significantly higher in ACC and ACC+LCA driving compared to manual driving. Similar to the 0-50 km/h condition the long glance probability during ACC+LCA driving is also significantly higher than during ACC driving.

Considering curved driving conditions, there is no significant difference in the long glance probability between manual and ACC driving across all speed conditions. However, for curved driving at speeds of 0-50 km/h we observe a significant increase in the long glance probability during ACC+LCA driving compared to manual driving (see Figure 8.5). For speeds of 50–100 km/h and speeds above 100 km/h, the increase in the long glance probability during ACC+LCA is significant compared to both, manual driving and ACC driving. For curved driving, no significant difference can be observed between manual driving and ACC driving. We can also observe that the confidence intervals for all ACC conditions are more widespread compared to the manual driving and ACC+LCA driving conditions.

Also shown in Figure 8.5, is the tendency that the long glance probability decreases with an increase in vehicle speed. This is in line with the model coefficients reported for Model 4 in Table 8.3. The same holds true for curved driving. Post-hoc pairwise comparisons show that during curved driving drivers' long glance probability decreases significantly across all conditions except ACC driving at speeds 0-50 km/h ($p = 0.5077$).

8.3 Discussion

In the following provide answers to our research questions and put our results into perspective.

8.3.1 The Effect of Driving Automation on Tactical Self-Regulation

Our findings on drivers' tactical self-regulation show that drivers adapt their interactions with the center stack touchscreen based on the automation level, vehicle speed, and road curvature (RQ1). Our results show that, drivers perform more touchscreen interactions per sequence with an increasing level of driving automation. Whereas speed influences the number of interactions only slightly, the number of interactions during curved driving decreases by 16 %. By breaking down interactions into specific touch gestures, we show that drivers in particular perform more complex gestures like drag and multitouch gestures during automated driving. Drivers also perform significantly less drag and multitouch gestures during curved driving compared to regular tap gestures. Both findings together suggest that drivers adapt their behavior to avoid complex touch gestures in demanding driving situations. They rather engage in such interactions during times of low driving demand. This is in line with the findings of Noble et al. [153] who found that drivers were more likely to perform high-risk secondary tasks during automated driving sequences.

Concerning drivers' interaction with specific UI elements, we show that during ACC+LCA driving, drivers interact particularly more often with lists or maps compared to other elements like the homebar or AppIcons. A potential explanation for this behavior is that lists and maps are visually more complex and drivers seem to perform these interactions in less demanding driving situations, e.g., with automation enabled or while driving straight. In contrast, the homebar, for example, is easy to access as it is visible on every screen and always located in the same position. The probability of drivers interacting with elements located at the homebar even decreases during ACC+LCA driving compared to manual driving. This could be due to many reasons. One of which may be that during situations of less driving demand, drivers prefer to use interfaces that allow for more control. For example, while it's possible to skip to the next song or radio station using the skip button on the homebar, the media app offers a complete overview of available songs and radio stations. Thus drivers have full control and can choose whatever they prefer. Whereas we observe similar trends for interactions during ACC driving, none of the differences proved to be significant. In earlier stages of this work [4], only using a subset of the data set, we found that these differences were statistically significant. This could be due to modifications in the UI software. Since the data is collected from test vehicles, the software is regularly updated so that the UI versions are optimized over time in terms of design, performance, and stability.

Considering drivers' behavioral adaption of touch gestures and UI elements, it is noticeable that drivers' self-regulation of complex interactions is more sensitive to changes in the driving demand than that of simpler interactions. Meaning that with an increasing driving demand the number of complex interactions decreases faster compared to simpler interactions and vice versa. These findings are in line with previous work [26, 144, 145, 146, 147], suggesting that drivers tend to perform more demanding tasks in less demanding driving situations. In contrast to related work, which mostly investigates the effects of drivers' tactical self-regulation on a task level, we show that these effects also exist on an interaction level. These new insights can help inform future UI designs for center stack touchscreens.

8.3.2 The Effect of Driving Automation on Operational Self-Regulation

In this study, we show that drivers not only adapt their glance behavior according to the level of driving automation (RQ2), vehicle speed, and road curvature but also show that significant interdependencies between these factors exist (RQ3). These novel findings suggest that drivers extend the margins to which they consider it safe to focus on the center stack touchscreen with an increasing level of driving automation. Even though drivers are supposed to constantly supervise the driving automation [30], the median glance duration during touchscreen interactions in ACC+LCA driving is 0.59 s longer than in manual driving. In comparison, Morando et al. [261] report an average increase of 0.3 s for glances to the center stack regardless of drivers interacting with the touchscreen. In line with the findings of Noble et al. [153], Gaspar and Carney [151], and Morando et al. [261], we also show that drivers are more likely to perform glances longer than two seconds when driving automation is enabled. Whereas Morando et al. [261] report an increase in the long glance probability toward the center stack touchscreen between manual and level 2 driving of 425 %, our results are similar to that of our previous study [4] and suggest an increase of 263 %. While the trend is similar, the absolute difference is probably due to differences in the driving contexts, the systems under test, or the data acquisition.

We also show that during ACC+LCA driving, drivers significantly increase in their mean glance duration toward the center stack touchscreen. This effect is statistically significant across all driving conditions and in line with the model explanations provided in Chapter 9. In contrast, Noble et al. [153] and Morando et al. [23] found no significant differences in the mean off-road glance duration for ACC or LCA driving compared to manual driving. There may be two reasons for this: First, the amount of data we leverage in this study is larger. Second, our eye tracker explicitly detects glances toward the center stack touchscreen that we then map to UI interactions. In other studies [22, 23, 152, 153], authors could not differentiate between general off-path glances, which might still be driving-related, and distraction-related off-path glances. This, inevitably, increases the number of false positives, making it harder to obtain significant results. Considering drivers self-regulation during ACC only driving, drivers increase their glance duration only for straight driving sequences and at speeds between 50–100 km/h and speeds above

100 km/h. For all other driving situation the effect is not significant. This suggests that drivers trust the ACC+LCA system to take over at least parts of the driving task in a wide variety of driving situations. On the other hand they only make use of the benefits of the ACC system in relatively controlled driving situations.

8.3.3 Limitations and Future Work

Naturalistic driving studies allow us to observe drivers in their natural driving environment. Driving simulator studies or test track studies, in contrast, suffer from an *instruction effect* because participants need to perform predefined tasks under specific conditions [142]. Furthermore, by leveraging production systems, we collect a large amount of data without the need for, potentially, error-prone manual labeling. However, certain limitations should be considered when interpreting the results.

All cars that contributed to the data collection are company internal test cars. Whereas, they are subject to various testing procedures but also for transfer and leisure rides of employees. Yet, the results of our data analysis do not indicate that specific UI stress tests have been conducted while driving. Furthermore, we argue that even during certain test protocols to evaluate driving-related functions, the incentive to interact with the IVISs does not differ from real-world driver behavior. Nonetheless, it is important to note that the software in these test cars is frequently updated and improved. This applies to the UI software as it does to the camera or ADAS software. This can lead to changes over time in the way drivers interact with the UI or how they self-regulate their behavior with regard to the driving demand. Compared to our previous work [4], we can observe differences in the glance and interaction behavior. The differences suggests that drivers' self-regulative behavior is sensitive to small changes in the UI or ADAS capabilities. To better understand this effect, similar naturalistic driving studies that compare various IVISs and ADASs are needed.

Another limitation that is that drivers need to be considered expert users. They are familiar with the cars and additionally obtained a prototype driver's license. Yet, the effect this might have is not clear. Whereas more experienced drivers tend to distribute their visual attention more adequately [262], Naujoks et al. [263] report that drivers who are familiar with driving assistance systems are more likely to engage in secondary tasks during assisted driving compared to drivers with no experience. In general, the glance duration distribution is roughly similar to those reported in related studies [23, 151, 153].

Due to data privacy regulations, we cannot differentiate between individual drivers. We can only differentiate between different trips and car types. Considering that more than 100 cars, with even more individual drivers, contributed to the data collection, the risk of overfitting to particular drivers is small. However, it is important to consider that only employees contributed to the data collection. For this reason, the results are likely biased toward mid-age drivers.

As we cannot differentiate between individual drivers, we are not able to show personal differences in drivers' self-regulative behavior. However, most of the models fitted (e.g., Model 2 and Model 4 in the supplementary material[4]) in this study have a significantly smaller Marginal R^2 compared to the Conditional R^2 [264]. This indicates that most of the covariance in the data is explained by the fixed and random effects together rather

[4] http://kups.ub.uni-koeln.de/id/eprint/65348

than by the fixed effects only. Even though we only incorporate the trip ($n = 10,402$) and car type ($n = 138$) as random effects these difference in the Marginal R^2 and Conditional R^2 suggest that trip-related or personal differences might influence self-regulation. This in line with previous research, but quantifying this effect based on naturalistic data could be the logical next step. The effect of task priority on self-regulation [265] is another factor that is not currently considered, but may provide insights that can aid the design of IVISs.

This work could be further improved by incorporating more features that describe the driving demand. Currently, we do not consider environmental factors such as weather and daylight. Speed and curvature may also not be sufficient to distinguish between different driving situations. Low speed and straight driving might be typical for traffic jam behavior (very controlled driving scenario), but also for city driving (very uncontrolled driving scenario). Including these features could help to provide a more holistic picture of drivers' behavioral adaptations to driving demands.

8.4 Conclusion

We present the first naturalistic driving study to investigate tactical and operational self-regulation of driver interactions with center stack touchscreens. Understanding self-regulation is key to understanding the effects of automation and assistance functions on driver distraction and driving safety. Furthermore, knowledge about self-regulation may help design more user-centered and context-aware IVISs. The key strengths of our study over the state-of-the-art are two-fold: (1) The large amount of naturalistic data, compared to related approaches [23, 153, 263], allows us to investigate drivers' tactical and operational self-regulation in greater detail concerning the driving demand. (2) We evaluate self-regulation specifically during interactions with the center stack touchscreen by combining driving data, UI interactions, touch gestures, and explicit glances toward the center stack touchscreen. That makes this the first naturalistic driving study to show self-regulation based on the analysis of touchscreen interactions.

Our modeling results show that driving automation has a stronger effect on self-regulation than vehicle speed or road curvature. Drivers interact more with the IVIS when ACC or ACC+LCA is enabled, use more complex UI elements, and perform more complex touch gestures. Even though driving assistance functions up to level 2 still demand the driver to have full control over the car, we observe 36% longer glances toward the center stack touchscreen when ACC+LCA is active.

Further research is needed, but based on the assumption that drivers kept the driving similarly safe throughout all conditions, fixed limits for acceptable demand as reported in the NHTSA Driver Distraction Guidelines [15] need to be adjusted according to different levels of driving automation and driving demands.

Chapter 9

Visual Demand Prediction of IVIS Interactions

Context One of the main principles of UCD is to design iteratively and with continuous validation. To achieve this, users must be involved throughout the design process. However, the increasing number of functions of IVISs and the different driving contexts that need to be considered when evaluating driver interactions with these systems make the design task more complex and user involvement more expensive. The results presented in Chapter 4 and Chapter 5 show that throughout the design-phase UX experts often struggle to make user-centered design decisions due to time restrictions. Thus, for most of the design phase, UX experts evaluate their designs in small-scale studies that do not replicate the driving context. Often, designs are only evaluated within the design team, and design decisions are made based on the experts' gut feeling. However, as shown in Chapter 8, drivers significantly adapt their interaction behavior according to variations in driving demands. Yet, OEMs only conduct resource-intensive empirical simulator studies once the design has reached a certain level of maturity. However, these tests represent only a fraction of the various facets of real-world driving. While the increasing design complexity makes the involvement of real users more and more expensive, feedback from simulated users can play a crucial role in the design phase. Embedded in the iterative design process, computational models of human-vehicle interaction can complement the capabilities of human UX experts [59]. They provide a cost-effective and dynamic method that can provide feedback throughout all stages off the design phase and can facilitate the development of interaction concepts that are safe by design and, therefore, less likely to fail final safety evaluations. This meets the needs of automotive UX experts for methods to automatically evaluate designs (see Chapter 4).

Contribution In this chapter, we showcase how to use machine learning methods to predict and explain the visual demand of in-vehicle touchscreen interactions based on large naturalistic driving data. The contribution of this work is two-fold: First, we propose a machine learning approach predicting the visual demand of in-vehicle touchscreen interactions based on the type of interaction and the associated driving context. Second, we apply the SHAP method [266] to explain the predictions and to visualize how user interactions, vehicle speed, steering wheel angle, and automation level, affect drivers' long glance probability and total glance duration. Overall, our method can build the basis for automated data-driven evaluations of early-stage IVIS prototypes.

Related Publications This chapter is adapted with minor changes from Ebel et al. [6].

9.1 Proposed Approach

The goal of this approach is to predict drivers' visual attention allocation based on user interactions and the associated driving parameters. To do so, we model drivers' secondary task engagements by combining interaction sequences, driving sequences, and glance sequences. These concepts are introduced in Section 2.4. In the following, we introduce the SHAP method and elaborate on the data processing and modeling procedure.

9.1.1 Explainable Predictions with SHAP

Explainable AI (XAI) aims to make machine learning models more transparent by providing human-understandable (interpretable) information, explaining the behavior and processes of machine learning models [267, 268]. Explanations can be understood as an interface between the human and the model [267] and can be valuable in various applications [269, 270]. Explanations can enhance scientific understanding [271], increase user trust [272, 273], and can be used to infer causal relations in data [274]. For the task at hand explainable predictions are of particular interest because the goal is not only to make predictions of the visual demand but also to draw conclusions about the impact of specific UI elements, gestures, and varying driving situations. The goal of this approach is to enable AI-assisted decision-making [275], optimizing a joint decision based on the domain knowledge of the human expert and the insights generated by the model prediction and accompanying explanation.

SHAP, proposed by Lundberg and Lee [266] is a method based on Shapley values from coalitional game theory [276]. The SHAP method provides local and global explanations for arbitrary predictive models. SHAP belongs to the class of additive feature attribution methods. The main idea is to use an interpretable explanation model $g(z')$ in the form of a linear function such that the model's prediction of a certain instance is equal to the sum of its feature contributions $\phi_i \in \mathbb{R}$ [277]:

$$g(z') = \phi_0 + \sum_{i=1}^{M} \phi_i z'_i \,, \tag{9.1}$$

where $z' \in \{0,1\}^M$ with z'_i represents the presence of feature i, ϕ_0 represents the models output in case no feature is present, and M is the number of input features [278].

Lundberg and Lee [266] further state that a single unique solution exists that follows the definition of additive feature attribution methods (see Equation 9.1) and satisfies the properties of local accuracy, missingness, and consistency. Local accuracy describes that the sum of the feature attributions is equal to the prediction of the original model. Missingness describes that a missing feature ($z_i = 0$) gets assigned an attribution of zero and consistency states that when changing a model such that it is more dependent on a certain feature, the attribution of that feature should not decrease.

The only possible solution as described by Lundberg and Lee [266] is given by the SHAP values:

$$\phi_i = \sum_{S \subseteq N \setminus \{i\}} \frac{|S|!(|M| - |S| - 1)!}{M!} \left(f_x(S \cup \{i\}) - f_x(S)\right) \,, \tag{9.2}$$

with S being the set of non-zero indexes in z', $f_x(S)$ being the expected value of the function conditioned on a subset S of the input features, and N being the set of all input features.

Multiple different approaches exist to approximate SHAP values for different kinds of machine learning models. However, in this study, we use TreeSHAP [278] which allows the computation of exact SHAP values for tree-based approaches. Compared to approaches like Local Interpretable Model-Agnostic Explanations (LIME) [279] or approaches specific to tree-based models like permutation importance of feature impurity calculations, SHAP has many advantages. Due to the solid foundation in game theory [277], SHAP values come with theoretical guarantees about consistency and local accuracy. Additionally, local and global explanations are consistent, SHAP values indicate whether the contribution of each feature is positive or negative, and Lundberg et al. [280] show a greater overlap of SHAP values and human intuition [266, 278].

9.1.2 Data Collection and Processing

The data used in this study was collected over the air from over 100 test vehicles and five different vehicle models using the Mercedes-Benz Telematics Data Logging Framework introduced in Section 3.1. The data collection period ranged from mid-October 2021 to mid-January 2022. We did not collect demographic or environmental data. The data was processed according to the procedure described in Section 3.2 and the extracted sequences follow the definitions given in Section 2.4.

In total, we extracted 322,425 touchscreen sequences. We obtained valid speed data for 145,973 sequences, valid steering data for 81,150 sequences, and valid glance data for 111,792 sequences. After individual processing, we computed the intersection of the individual data sources resulting in 30,158 complete secondary task engagements. Most of the sequences were excluded either because they were generated on a test bench (no driving data was available), the car wasn't equipped with a camera, or because the sampling requirements were violated due to a loss of data connection. In the second stage of data processing, we apply further filtering steps to increase data quality. To prevent the data from becoming too sparse, we discarded 342 secondary task engagements with more than 41 interactions ($N > 41$ corresponds to the 99^{th} percentile of the distribution of N). These secondary task engagements can be considered outliers without providing additional benefits for the use case at hand. We further discard 16,864 secondary task engagements where a passenger was present because they also tend to interact with the center stack touchscreen. These interactions can not be mapped to driver glances and would skew the data toward fewer and shorter glances per secondary task engagement with many interactions logged that did not originate from the driver. Furthermore, we discarded 809 engagements during which the car came to a full stop and one sequence due to a remaining speed error. After this processing step, we obtained the final set of 12,142 secondary task engagements. Finally, we compute summary statistics for the secondary task engagements to generate the final set of features as described in Table 9.1. These features serve as input to the models introduced in the following.

Table 9.1: Overview of the final input features that describe a secondary task engagement.

Feature	**Description**
Interaction Data	
n_{Button}	# Interactions with regular buttons (e.g., push or radio buttons)
n_{List}	# Interactions with lists (e.g., choosing a suggested destination)
n_{Map}	# Interactions with a map viewer (e.g., interacting with the navigation map)
n_{Slider}	# Interactions with slider elements (e.g., changing the volume)
n_{Homebar}	# Interactions with the static homebar on the bottom of the screen
$n_{\text{CoverFlow}}$	# Interactions with cover flow widgets (e.g., scrolling through albums covers)
n_{AppIcon}	# Interactions with app icons on the home screen
n_{Tab}	# Interactions with tab bars
n_{Keyboard}	# Interactions on the keyboard or number pad (e.g., entering a destination)
n_{Browser}	# Interactions within the web browser
n_{RemoteUI}	# Interactions within Apple Car Play or Android Auto
$n_{\text{ControlBar}}$	# Interactions with a control bar, displayed as a small overlay
n_{PopUp}	# Interactions with pop-up element
$n_{\text{ClickGuard}}$	# Interactions with non-interactive background elements
n_{Other}	# Interactions with a UI element that does not fit any of the categories
n_{Unknown}	# Interactions with a UI element for which no identifier exists
n_{Tap}	# Tap gestures
n_{Drag}	# Drag gestures
$n_{\text{Multitouch}}$	# Multitouch gestures
d_{avg}	Average distance between two consecutive interactions in px
N	Number of interactions
Driving Data	
v_{avg}	Average vehicle speed in km/h
θ_{avg}	Average steering wheel angle in °
a_{acc}	Status of the adaptive cruise control $a_{\text{acc}} \in \{0, 1\}$
a_{lca}	Status of the steering assist $a_{\text{lca}} \in \{0, 1\}$

9.1.3 Modeling

As formulated in the problem statement we solve one classification task (long glance prediction) and one regression task (total glance duration prediction). For each of the tasks, we compare a *Baseline* approach and a *Logisitc/Linear Regression* approach against three machine learning approaches, namely *Random Forests*, *Gradient Boosting Trees*, and *Feedforward Neural Networks (FNNs)*. In the long glance prediction task, the *Baseline* approach randomly predicts one of the two classes (i.e. in a balanced dataset the probability of correctly predicting a long glance is roughly 50%). In the total glance duration prediction task, the baseline approach predicts the median total glance duration of the training dataset. The parameters of the machine learning-based methods are chosen based on extensive hyperparameter optimization using random search.[1].

9.2 Evaluation

In this section we present the final dataset, put the experimental results in perspective, and elaborate on the explainable predictions generated by applying the SHAP method.

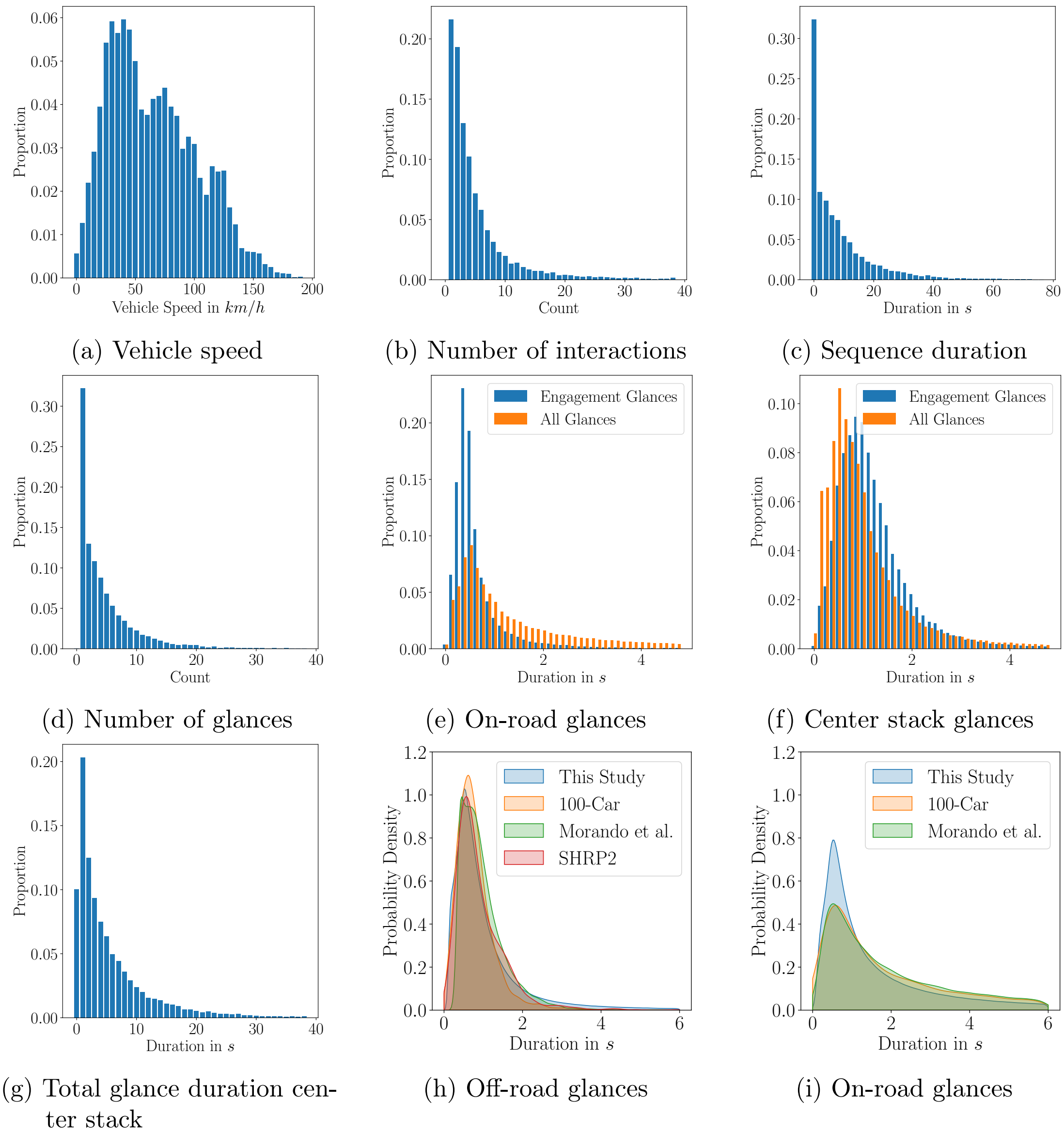
9.2.1 Dataset

The final dataset consists of 12,142 secondary task engagements sampled from 3,046 individual trips. The majority of secondary task engagements were collected from the Mercedes-Benz S-Class (7,342 secondary task engagements), EQS (3604), and EQE (824) models. The cars were equipped with a 17.7", 12.8", or 11.9" center stack touchscreen with similar pixel density. In total, 61,943 touch interactions and 119,770 individual glances were collected. The median trip length is 34.28 minutes ($Q_1 = 17.49, Q_3 = 66.58$). Specific glance and interaction statistics of the final dataset are presented in Figure 9.1.[2]

In Figure 9.1e and Figure 9.1f, the glance duration distribution during secondary task engagements (blue) is plotted against the glance duration distribution over all sessions independent of the driver being engaged in a secondary task (orange). This allows a comparison with approaches that utilize data collected irrespective if the driver being engaged in a secondary task or not.

We further compare our data with the manual driving baseline of the 100-Car Study [78] (data provided by Custer [281], the SHRP2 [79] (data available in Bärgman et al. [282]) and the data reported in the work of Morando et al. [23] (provided by the authors upon request). Figure 9.1i and Figure 9.1h show the glance distribution of on-road and off-road glances for the respective datasets. The glance distributions were truncated at 6 seconds since this corresponds to the length of the segments in the 100-car baseline dataset. The visual comparison shows that the off-road glance duration distribution matches well with the data reported in the three related studies. However, the on-road glances show some differences between our data and the data reported in the 100-Car study and the study of Morando et al. [23]. Whereas the mode is similar for all three datasets, the on-road glances in our study tend to be shorter compared to the other two studies. The potential reasons for this are manifold. For example, Morando et al. [23] only consider driving segments of very controlled driving by excluding curved driving, lane changes, and driving segments

[1]For more details on the search space refer to: `http://kups.ub.uni-koeln.de/id/eprint/65348`
[2]For the full dataset statistics refer to: `http://kups.ub.uni-koeln.de/id/eprint/65348`



(a) Vehicle speed (b) Number of interactions (c) Sequence duration

(d) Number of glances (e) On-road glances (f) Center stack glances

(g) Total glance duration center stack (h) Off-road glances (i) On-road glances

Figure 9.1: Descriptive statistics of the glance and interaction data. The histograms show the (a) average speed per secondary task engagement, (b) the number of interactions per secondary task engagement, (c) the duration of the interaction sequences of a secondary task engagement, and the (d) number of glances toward the touchscreen during per secondary task engagement. Figure (e) compares the on-road glance duration distribution for glances during a secondary task engagement with glances irrespective whether a touchscreen interaction was performed or not. Figure (f) establishes the same comparison for glances toward the center stack touchscreen. Figure (g) shows the total glance duration toward the center stack touchscreen during a secondary task engagement. Figure (h) and (i) compare the probability density functions of the off-road and on-road glance duration from this study with the 100-Car study, the SHRP2 study (only off-road glances), and the study of Morando et al. [23].

Table 9.2: Comparison of the different models. Accuracy and Mean Absolute Error (MAE) are given together with the Standard Deviation (SD).

	Long Glance Prediction		Total Glance Duration Prediction	
Model	Accuracy	SD	MAE	SD
Baseline	50.09 %	1.63 %	4378 ms	177 ms
Logistic/Linear Regression	61.93 %	1.69 %	3778 ms	383 ms
Random Forest	**67.53 %**	**1.38 %**	**2437 ms**	**112 ms**
XGBoost	67.22 %	1.85 %	2385 ms	117 ms
FNN	65.90 %	2.02 %	2443 ms	109 ms

with a vehicle speed under 60 km/h. In these rather calm driving situations, drivers need to switch less often between the road and off-road regions such as mirrors or side windows resulting in longer continuous on-road glances. The differences with regard to the 100-Car study could be due to the fact that the data is now almost 20 years old and only covers manual driving. The technology of the vehicles at that time, and in particular that of the infotainment and assistance systems, was fundamentally different from that in today's vehicles. However, considering the differences in the data collection, the comparison suggests that our data collection and processing pipeline produces representative data.

Figure 9.1e indicates that during touchscreen] interactions, drivers need to distribute their visual attention between the road and the center stack resulting to shorter on-road glances. On the other hand, center stack glances during secondary task engagements tend to be longer than general center stack glances (Figure 9.1f). Through Figure 9.1b, we see that roughly 25 % of all sequences consist of only a single interaction. This results in many short secondary task engagements that only consist of a single glance toward the center stack (Figure 9.1d). These short engagements are part of real-world user behavior. However, they are often not represented in laboratory studies where only a few predefined tasks are evaluated. We argue that it is still relevant to analyze these short engagements and therefore decide to consider them. For the long glance classification task, we balanced the dataset by applying random undersampling. The resulting dataset consists of 4,816 sequences for each class.

9.2.2 Experimental Results

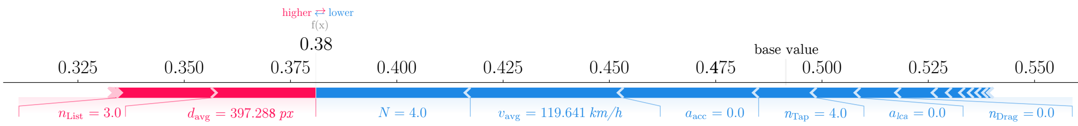
We evaluate the regression models using a repeated 10-fold cross-validation [283] and the classification models using a stratified 10-fold cross-validation. The results are given in Table 9.2. The models were fitted on the full set of input features given in Table 9.1.

The machine learning-based approaches outperform the Baseline approach and the Logistic and Linear Regression approaches in both tasks. The differences in the prediction accuracy support our assumption that neither of the problems at hand can be considered a linear problem and that interaction effects between different features exist. The machine learning models provide similar results. However, the Random Forest approaches offer two desirable properties making them in particular suitable for the use case at hand. First, the TreeSHAP [278] algorithm allows efficient computation of exact SHAP values for Random Forest models. Second, Random Forests can be run in parallel, making them suitable for future use cases when they are deployed on data of a whole production fleet. Thus, we choose the Random Forest models for the following explanation generation.

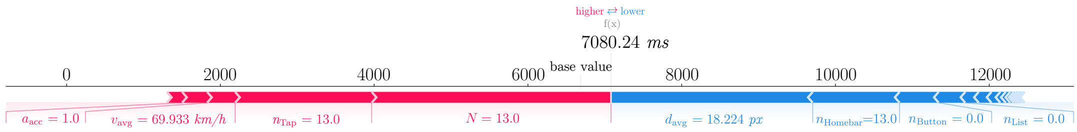
9.2.3 Explainable Predictions

While the above-presented results provide a good measure of prediction accuracy, they are of limited value when it comes to understanding human behavior. To truly support researchers and practitioners in the design process to foster a deeper understanding of drivers' visual attention allocation, it needs more than just predicting whether a new user flow might cause too much distraction [2]. For this reason, we employ SHAP. SHAP values represent the features' contribution to the model's output, providing a local explanation for each input sample. By combining many local explanations, one can represent global structures producing detailed insights into model behavior [278].

Local Explanations Figure 9.2 displays the explanations for one long glance prediction and one total glance duration prediction. These force plots represent a particular model output as a cumulative effect of feature contributions (i.e. SHAP values). The length of each bar indicates how much the associated feature value pushes the model output from the base value toward higher values (red, to the right) or lower values (blue, to the left). The base value is computed as the average model output over the training dataset. The features in each group are sorted based on the magnitude of their impact and only the most influential features are displayed. The feature values are shown below the bars. For the long glance prediction, feature contributions are displayed as probabilities. For the total glance duration prediction, they are shown in milliseconds.



(a) Long glance prediction sample



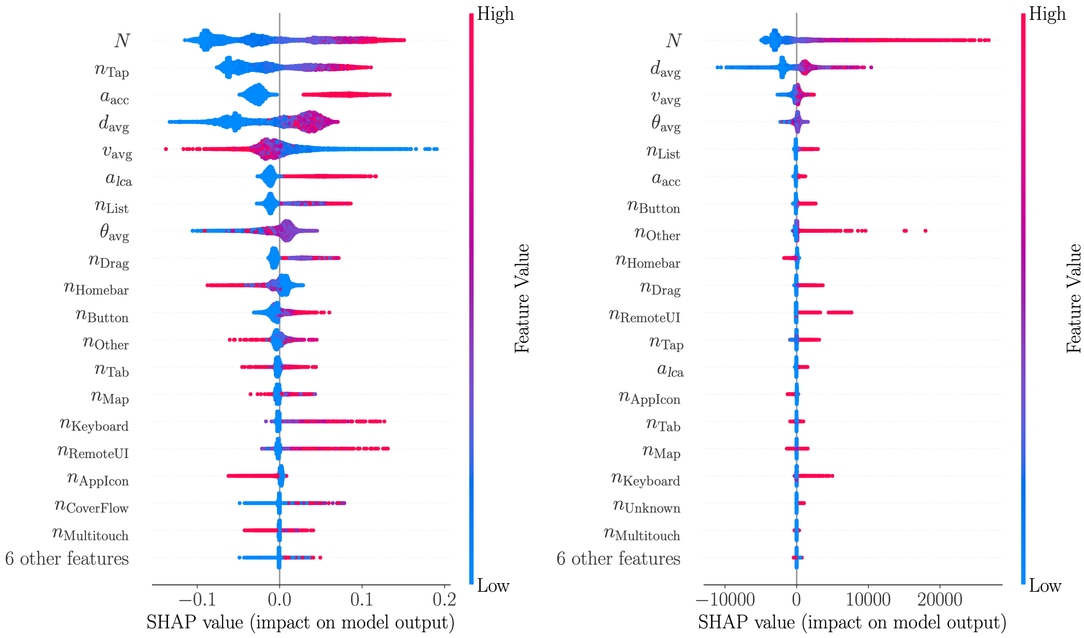
(b) Total glance duration prediction sample

Figure 9.2: Local explanations of a specific secondary task engagement visualized as force plots.

Figure 9.2a visualizes the explanation of a secondary task engagement for which the model outputs a long glance probability of $0.38 = 38\,\%$. The long glance probability is pushed to the left because the driver only performed 4 interactions ($N = 4$) and drove at a speed of $v_{\text{avg}} = 119.641\,km/h$ while the ACC was deactivated ($a_{\text{acc}} = 0$). On the other hand, the prediction is pushed to the right because the interactions were quite distributed over the screen ($d_{\text{avg}} = 397.288\,px$) and three of them were list interactions ($n_{\text{List}} = 3$).

Another secondary task engagement is explained in Figure 9.2b. Here, the total glance duration prediction of roughly $7\,s$ is close to the base value because the positive and negative feature contributions balance each other out. During this secondary task engagement, the driver performed 13 touch interactions ($N = n_{\text{Tap}} = 13$) while driving with an active ACC ($a_{\text{acc}} = 1$) at a speed of 70 km/h. If the model would only access this information, it would predict a total glance duration of roughly 13 seconds However, as all interactions were very close to each other ($d_{\text{avg}} = 18.224\,px$) and were all performed on the homebar ($n_{\text{Homebar}} = 13$ without any list or button interaction interfering ($n_{\text{List}} = 0$, $n_{\text{Button}} = 0$, the final model output is only slightly higher than the average total glance duration prediction.

These local explanations show that not all features are always relevant. Predictions for secondary task engagements can be driven by only a few dominant features. The presented explanations enable designers and researchers to quickly identify the main forces behind individual predictions. It also allows them to play around with artificial input samples and observe how certain changes in the design of a user flow or the driving situation impact the model's output.



(a) Long glance prediction

(b) Total glance duration prediction

Figure 9.3: Explanation summary visualized as a set of beeswarm plots. Each beeswarm plot represents the distribution of SHAP values for one feature.

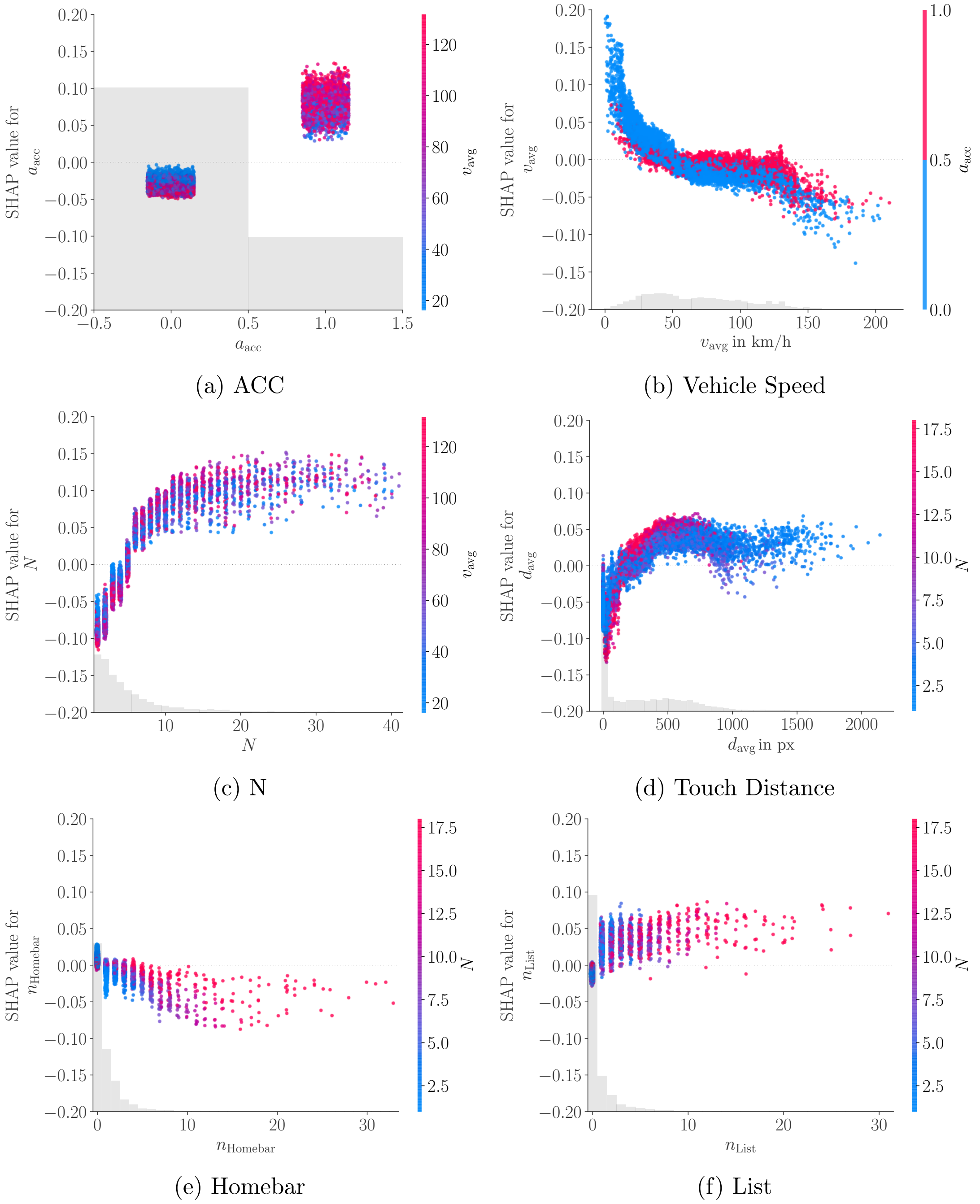
Global Explanations To understand how the features affect the model's output on a global scale, we combine all local explanations of the dataset. Figure 9.3 shows the distribution of SHAP values (i.e., the impact of each feature on a specific prediction as seen in Figure 9.2) as a set of beeswarm plots. Each dot in a row corresponds to an individual secondary task engagement. The position on the x-axis represents the effect of the respective feature on the model's output. In Figure 9.3a, the SHAP values are in probability space, and in Figure 9.3b they represent the impact in milliseconds. The color indicates the feature value (red is high, blue is low). The features are sorted by their global importance and only the 19 most important features are displayed individually.

The most important features of the long glance prediction model (Figure 9.3a), are the number of interactions N, the average distance between the interactions d_{avg}, and the number of tap gestures n_{Tap}. The more touchscreen interactions a driver performs and the larger the distance between them, the higher the output probability that one of the associated glances is longer than 2 seconds. Figure 9.3a also reveals that both, the activation of ACC a_{acc} and LCA a_{lca}, increase the long glance probability. Whereas the impact of a deactivated assistance system (blue) is small for all samples, the impact varies if the assistance systems are active. The horizontal spread suggests that the impact of assisted driving on visual attention allocation is situation-specific and depends on further factors like the driving situation and interaction patterns. The distribution that describes the impact of the vehicle speed v_{avg} is heavily tailed. For most secondary task engagements at medium speed, the effect is negative but rather small. High speed values reduce the predicted long glance probability and low speed values increase it, respectively. This indicates drivers' self-regulative behavior.

The number of list interactions n_{List} is the most important feature associated with a specific UI element followed by the number of interactions with the homebar n_{Homebar}. Through Figure 9.3a, we see that their impact is opposite to each other. Whereas the long glance probability increases with an increasing number of list interactions, it decreases for an increasing number of homebar interactions. This suggests that list interactions tend to be more distracting than interactions on the static homebar. The impact of interactions with Android Auto or Apple Car Play n_{RemoteUI} is similar to the impact of list interactions. In general, we can observe that most of the SHAP value distributions associated with a specific class of UI elements are centered around zero with long tails to one or both sides. This is because most of the elements occur in only a small portion of secondary task engagements. Whereas this leads to a relatively low global importance, these features still have a large impact on specific predictions.

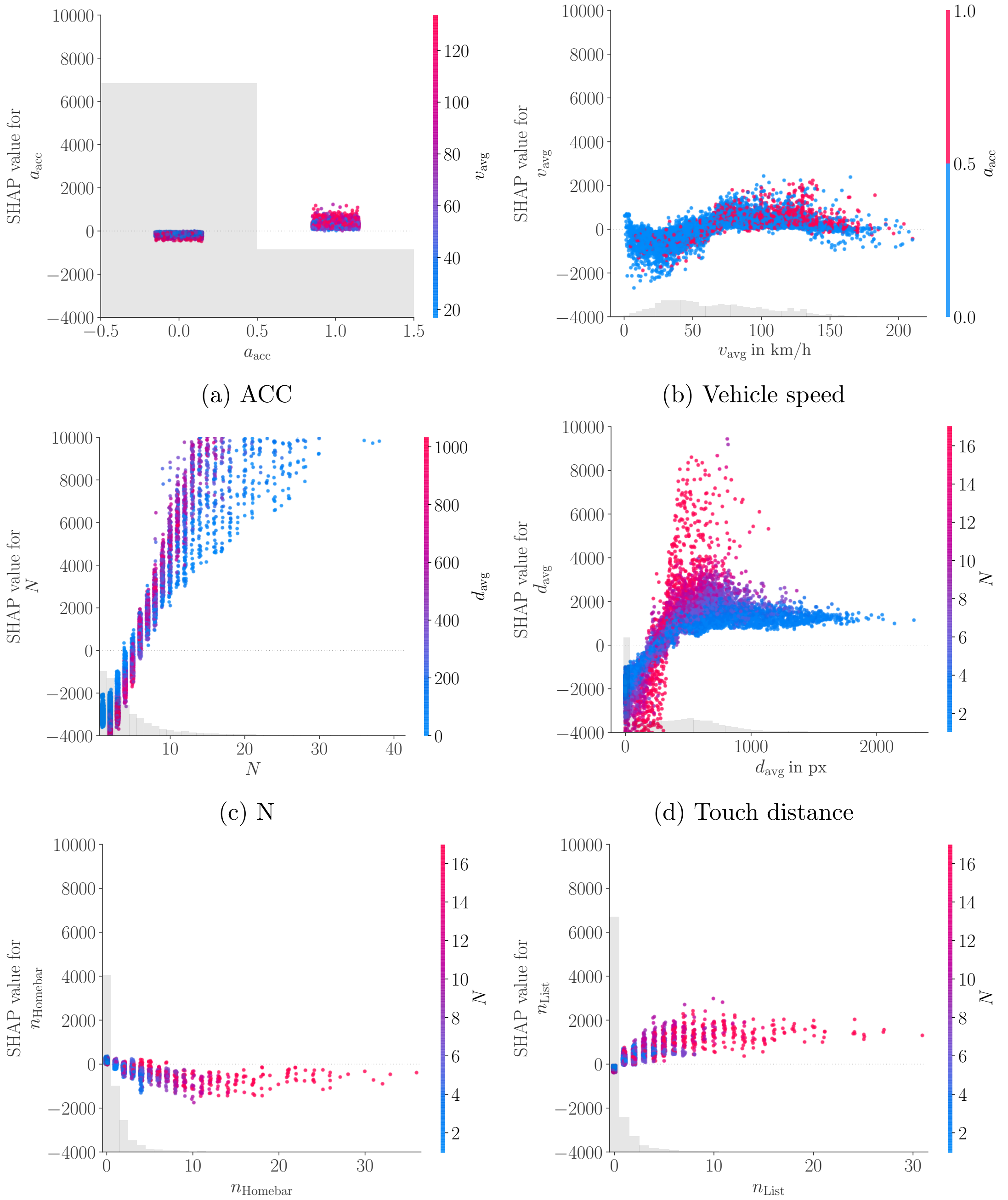
For the total glance duration prediction model (Figure 9.3b), N and d_{avg} are also the two most important features. Their distributions also show similarities to the distributions observed in the long glance prediction task. However, the impact of the vehicle speed v_{avg} is inverse compared to the long glance prediction task. High speed values increase the total glance duration prediction and low values decrease the prediction. Both findings together could be an indication that drivers reduce their single glance duration at higher speeds, which in turn results in longer total glance durations because more individual glances are required to complete the same task.

Further, we can see that there are almost no negative contributions associated with UI interaction features. This is due to the fact that the total glance duration prediction task is cumulative, and every interaction inevitably implies a certain amount of visual attention. However, homebar interactions n_{Homebar}, can negatively affect the model output. In line with the observations made for the long glance prediction, list interactions n_{List}, map interactions n_{Map} and interactions with Android Auto and Apple Car Play n_{RemoteUI} can be associated with an increased visual demand prediction. A comparison between Figure 9.3a and Figure 9.3b also reveals that the total glance duration is not as dependent on the status of the driver assistance systems as the long glance probability.



(a) ACC (b) Vehicle Speed (c) N (d) Touch Distance (e) Homebar (f) List

Figure 9.4: Feature dependence plots for the long glance classification model.



(a) ACC (b) Vehicle speed

(c) N (d) Touch distance

(e) Homebar (f) List

Figure 9.5: Feature dependence plots for the total glance duration model.

To understand the effect of a single feature on the model's output in more detail, we plot the SHAP values (y-axis) against the corresponding feature values (x-axis). Every secondary task engagement in our dataset is represented as a dot (see Figure 9.4 and Figure 9.5). Vertical dispersion at a single value on the x-axis shows that there are non-linear dependencies between the displayed feature and other features. To highlight the interaction between features, each dot is colored by the value of the feature that shows the strongest interaction. The histogram at the bottom of the plots shows the distribution of datapoints. Figure 9.4a suggests that the use of ACC leads to an increased long glance probability prediction. The interaction with the vehicle speed shows that the effect tends to increase with increasing vehicle speed. On the other hand, the data shown in Figure 9.4b indicates that drivers tend to increase their single glance durations at lower speeds (below 50 km/h) and decrease them at higher speeds (above 125 km/h). However, in between those values, the speed has almost no influence on the model output. This suggests that drivers self-regulate their visual attention allocation based on what they consider an appropriate speed. Additionally, the interaction with the ACC status shows that the impact of the speed on the model output decreases when ACC is active. The interaction effect with a_{acc} partially explains the variance (vertical diversion) in the effect of the vehicle speed. However, various factors like road type or speed limit that may also influence how the vehicle speed affects drivers' visual attention allocation are not considered in the presented models.

Figure 9.4c indicates that the number of interactions is positively correlated with the drivers' probability to perform a long glance. On the other hand, Figure 9.4d suggests that as soon as the distance between the touch interactions exceeds a certain threshold (roughly 200 px), the effect on the long glance probability remains constant. Whereas homebar interactions decrease the probability of the model predicting a long glance (Figure 9.4e), list interactions (Figure 9.4f) push the model toward predicting a long glance. In addition, the interaction effect with the number of interactions indicates that the effect of both elements increases as their proportion within a sequence increases.

Figure 9.5 visualizes how the different features affect the total glance duration prediction. While the number of interactions N (Figure 9.5c) is the dominant feature it is also the feature with the highest interaction effect on all other features. Compared to Figure 9.4 and in line with the observations we made in Figure 9.3, we see that the ACC status a_{acc} (Figure 9.5a) and the vehicle speed v_{avg} (Figure 9.5b) do not influence the total glance duration prediction as much as they influence the long glance prediction. This applies in particular to secondary task engagements with few interactions. The impact of list interactions n_{List} and homebar interactions n_{Homebar} on the total glance duration, however, is similar to the impact those interactions have on the long glance probability (Figure 9.5e and Figure 9.5f). This also applies to the influence of the average touch distance d_{avg}. An increase in touch distance leads to an increase in total glance duration and long glance probability until a certain threshold is reached. However, the interaction effect with N is higher for the total glance duration prediction model. Another interesting aspect that might need further exploration is the location of the x-intercept. This point describes the touch distance at which the feature's impact turns from decreasing to increasing the visual demand prediction (Figure 9.5d).

9.3 Discussion

The presented approach enables users to evaluate the visual demand of early-stage prototypes. In the following, we put our results into perspective and show that the presented approach is more accurate than comparable methods. The predictions and explanations facilitate the generation of fast insights without requiring expensive and long-planned user studies. We illustrate this by assessing three exemplary research objectives covered in the literature. Finally, we address several limitations that apply to our approach.

9.3.1 Predicting the Visual Demand of In-Vehicle Touchscreen Interactions

Given the complexity of the modeling task, the presented results show how machine learning methods can be used to generate valuable insights into drivers' multitasking behavior by leveraging large naturalistic driving data. Compared to the approach of Kujala and Salvucci [175], who report critical differences between model predictions and observations, our approach is not only more accurate but also considers a more diverse set of UI elements.

Our approach can predict the total glance duration with a mean absolute error of roughly 2.4 seconds over a diverse range of interactions and driving scenarios. In comparison, Purucker et al. [170] report a mean error of 4 s when averaged over all evaluated tasks. Furthermore, Purucker et al. [170] use a simple car following task at a constant speed for evaluation. Although these comparisons are useful to put the results into perspective, one needs to consider that the approaches highly differ in their environments and scenarios as described by Janssen et al. [104].

9.3.2 Fast and Easily Accessible Insights

Our approach has two main advantages over conventional user studies. First, the models allow making predictions for yet unseen secondary task engagements. Conventional studies can only be used to evaluate situations that were explicitly tested. Second, if the user interface undergoes disruptive changes (e.g., a completely new design or concept), the results of a user study are no longer valid and a new study needs to be conducted. Similarly, our computational models may also lose their capability to generalize. However, the advantage of the automated approach for data collection and modeling is that, as soon as a new version is deployed to test vehicles, data is collected and new models based on this new version can be fitted. To demonstrate that our approach is a meaningful extension of traditional user research methods, we compare our results with those from conventional user studies.

The Influence of Vehicle Speed on Drivers' Visual Attention Allocation Based on data from the SHRP2 dataset, Risteska et al. [22] found that an increase in speed can be associated with a reduction in drivers' long off-path glances. They argue that drivers modulate their visual attention allocation based on driving demands. A similar finding is presented by Tivesten and Dozza [148], who found a significant correlation between vehicle speed and off-road glance duration when drivers were engaged in a visual manual phone task. This is consistent with our results shown in Figure 9.4b and Figure 9.3a. Our explanations do not only indicate that the long glance probability decreases with increasing speed but further suggest that this behavior might not be strictly proportional and is also affected by the status of driver assistance systems. Our results further show that the predicted total glance duration increases with increasing speed (see Figure 9.5b). The combination of both findings provides a more comprehensive picture suggesting that drivers reduce their single glance duration at higher speeds, forcing them to look to the center stack touchscreen more often. This, in turn, leads to an increased total glance duration because certain aspects of human glance behavior like the time needed to locate an item are constant for each glance [171].

The Influence of Driving Automation on Drivers' Visual Attention Allocation Assisted driving is associated with an increase in the mean and total glance duration during secondary task engagements [172, 284]. This is in line with our findings presented in Figure 9.3. In a driving simulator study, Carsten et al. [284] also found that the effect of lateral control (LCA) on driver engagement is larger than the effect of longitudinal control (ACC). Based on our data, we cannot confirm this finding. The reasons for this can be manifold but may well be due to the difference between real data and simulation data. Our results further differ from those of Morando et al. [23], who report no differences in the aggregate off-path glance duration distributions between manual and assisted driving. They only report an effect concerning the on-road glance distribution but state that their eye-tracker did not provide detailed information about the off-path AOIs. Since we can explicitly detect glances toward the center stack touchscreen and can distinguish them from general off-path glances, we argue that our results are superior.

The Influence of Design Characteristics on Drivers' Visual Attention Allocation There are not yet many approaches that have investigated the influence of design characteristics on visual demand in such detail (element type basis) as we show in our approach. Kujala and Salvucci [175] found that the average distance between two consecutive touch interactions is a critical factor associated with long glances exceeding the limit considered safe. This is in line with our results presented in Figure 9.4d and Figure 9.5d. The explanations that our method provides could additionally serve as a first attempt to quantify the impact spatial separation of interaction elements has on visual demand while driving. Our approach also allows us to make detailed statements about the influence of individual elements. So far, only the task interaction times have been studied in the literature in a roughly similar level of detail [165, 229]. We found that in particular interactions with maps, lists, and interactions within Apple Car Play and Android Auto seem to be visually demanding. Interactions on the static homebar, with app icons, and general buttons, on the other hand, are less demanding.

9.3.3 Benefits for the Design Process of IVISs and Implications on Distracted Driving Prevention

To develop IVISs that are safe to use, driver distraction evaluation needs to be an integral part already in the early design stages. However, driver distraction is a complex construct, and automotive UX experts need data-driven support to evaluate and compare design alternatives concerning their distraction potential [2]. Thus, our approach aims to inform the design process of IVISs from the bottom up to develop solutions that are the least distracting and safe by design. We envision our method to be used to dynamically evaluate early-stage IVIS designs. Users can assess hypothetical IVIS designs concerning their distraction potential in terms of visual demand. They can play around with artificial input samples to learn how changes in the user flow or driving scenario affect drivers' visual attention allocation. Our method then explains how each parameter contributes to the overall prediction. Thus, designers can better understand the effects of various UI elements, driving automation, and vehicle speed on driver distraction. This information can then be used to design IVISs that are less distracting and reduce the risk of accidents. The improved accuracy over comparable approaches and the three application examples show that our approach can make a major contribution to better understanding the complex construct of driver distraction and drivers' visual attention allocation during secondary touchscreen tasks.

9.3.4 Limitations and Future Work

As we leverage already commercialized technologies of Mercedes-Benz, we collected a large amount of behavioral data. We observed drivers' natural interaction behavior without explicitly telling them which touchscreen interactions to perform and therefore eliminate the so-called instruction effect [142]. While this approach has many advantages, especially over simulator and test track studies, several limitations apply. These limitations and their potential implications are discussed in the following.

Only company internal cars contributed to the data collection. Whereas they are used for a diverse range of testing procedures, they are also used for transfer and leisure rides of employees, for example over the weekend. We argue that, even if drivers follow a test protocol that aims to evaluate driving-related functions, the incentive to interact with the IVIS does not deviate much from real-world behavior. Furthermore, all drivers in this study need to be considered expert users. However, it is not yet entirely clear to what extent the gaze behavior of experts differs from regular users. Whereas Wikman et al. [262] report that experienced drivers allocated their visual attention more adequately [262], Naujoks et al. [263] show that experienced users of ADAS tend to increase their secondary task engagements compared to novice users. However, a comparison with related approaches [23, 151, 153] shows high agreement in total and average glance behavior. Still, the restricted sample of drivers and the fact they were driving alone, need to be considered when interpreting the results.

It is important to consider that the features used in this work do not capture all factors that influence drivers' visual attention allocation. In this study, we only consider the level of driving automation, vehicle speed, and the steering wheel angle to describe the driving situation. These features and their interactions provide valuable information (compare Figures 9.4a, 9.4b, 9.5b, 9.5a, and the steering wheel feature dependence plot in the supplementary material[3]), but they do not allow for a comprehensive description of the driving situation. For example, the effect of vehicle speed may vary not only based on the level of driving automation, but also on the type of road and traffic situation. Therefore, including additional features may not only improve the description of the driving situation but also make the existing features more meaningful by considering their interaction effects.

[3] http://kups.ub.uni-koeln.de/id/eprint/65348

Furthermore, it is important to put the results into context and to elaborate on the practical implications this might have. As demonstrated, the approach provides valuable insights into how design artifacts and environmental factors affect drivers' visual attention allocation. The predictions and explanations can guide designers to create interfaces that are less distracting and safer to use. However, even though our approach is superior to related approaches, it is not yet accurate enough to make pixel-precise predictions or to differentiate between minor changes in the driving environment (e.g., driving at 72 km/h vs. 75 km/h). To reliably evaluate the effect of such slight changes or to even act as a basis for driver distraction guidelines, the accuracy needs to be increased. Furthermore, we do not consider environmental factors like lightning conditions or street type (e.g., rural road or highway) or UI artifacts like element color and size that might also influence visual attention allocation. Including such features would provide a more holistic picture and probably more accurate predictions. Moreover, drivers tend to self-regulate their willingness to engage in secondary tasks based on the driving task demands [138, 139]. As a result, some interactions occur less frequently in certain driving situations, leading to fewer training data. Therefore, it is likely that prediction accuracy varies across driving situations.

The presented explanations do not imply causality, and therefore do not represent a complete assessment of drivers' visual attention allocation while being engaged in a secondary touchscreen task. However, the explanations help designers to identify the most informative relationships between input features and model outputs, which assist them in understanding the visual demand predicted by the machine learning model.

Having shown that this method delivers promising results, the main goal of future iterations is to improve prediction accuracy. First, a more holistic description of the driving situation by providing additional features like lighting conditions, the proximity of surrounding road users, or map data might lead to significant improvements. Second, considering user demographics like age or driving experience might also lead to better accuracy. Finally, a larger dataset is not only likely to benefit the algorithms presented in this work, but would also enable more sophisticated approaches like recurrent neural networks that can capture sequential information embedded in the interaction sequences.

9.4 Conclusion

In this chapter, we propose a machine learning approach that predicts the visual demand of secondary touchscreen interactions while driving, according to the type of interactions that are performed and the associated driving parameters. Our approach generates local and global explanations providing insights how design artifacts and driving parameters affect drivers' visual attention allocation. We evaluate the approach on a real-world driving dataset consisting of 12,142 secondary task engagements. Our best model identifies secondary task engagements during which drivers perform a long glance with 68 % accuracy and predicts the total glance duration with a mean deviation of 2.4 s. The analysis of the generated explanations reveals clear differences between the visual demand of specific touchscreen interactions and shows that drivers' visual attention allocation depends on the driving situation. In line with related research [22, 148], we show that drivers modulate their visual attention allocation based on the vehicle speed and the level of driving automation.

Our key contributions address many points that previous approaches [22, 79, 171, 285] have identified as desirable: (1) The approach leverages continuously collected large-scale real-world data providing realistic predictions of drivers' visual attention allocation during secondary task engagements. (2) The approach can easily be adjusted to incorporate additional features and to predict various metrics in addition to total glance duration and long glance probability (e.g., number of glances, total eyes off-road time, mean glance duration). (3) The local and global explanations provide detailed insight into the impact design artifacts and scenario parameters have on driver distraction prediction. (4) The approach can inform designers about potential implications their design may have and can guide them to design in-vehicle touchscreen interfaces that are safe to use.

Chapter 10

Conclusions, Limitations, and Outlook

This chapter concludes the thesis, revisits its contributions, and provides future research directions. In Section 10.1, we summarize our findings and provide answers to the three main research questions. We also discuss how the individual contributions relate to each other and, as an overall construct, can improve the automotive UX design process. In Section 10.2 we discuss the limitations of this work and provide an outlook on how the contributions can guide future research and influence industrial applications.

10.1 Conclusions

All of the contributions in this thesis focus on the problem that many decisions in the automotive UX design process are not user-centered due to a lack of customer insight. This can result in distracting user interfaces that do not meet customer needs (see Section 1.2). The four main contributions of this work address this problem by focusing on the three complementary research questions introduced in Section 1.3. As illustrated in Figure 10.1, these research questions also serve as the leading objectives of the respective chapters.

We classify our contributions according to the contribution types in HCI proposed by Wobbrock and Kientz [286]. Accordingly, we provide two *empirical* contributions (contributions 1 and 3), one artifact contribution (consisting of contributions 2.1 and 2.2), and one *methodological* contribution (contribution 4).

Part I consists of two empirical studies [1, 2] that form the basis of this thesis. Together, they answer **RQ1** and provide an understanding of how data-driven methods can facilitate the design and evaluation of IVISs and how they can be integrated into the UX design process. We provide an overview of the current challenges and specific needs of UX experts in the automotive UX design process. As a result, we present potential application areas where data-driven methods can be used to make the design and evaluation of IVISs more user-centered.

The contributions made in Part II and Part III provide solutions specifically developed according to the potentials identified in Part I (see Figure 10.1). In Part II, we present two *artifact* contributions. The visualizations (contribution 2.1) and the interactive analysis tool (contribution 2.2) together present an answer to **RQ2**. In Part III, we answer **RQ3** by providing insights from an *empirical* study (contribution 3) and by presenting a method to predict the visual demand of in-vehicle touchscreen interactions. Thus, contribution 4 is considered *methodological*. However, the distinction between the contribution types according to Wobbrock and Kientz [286] is not always obvious. For example, it can be argued that Chapter 6, in addition to its artifact contribution, also provides a methodological component in the way the visualizations are generated and inform the design process.

Following the taxonomy proposed by Stol and Fitzgerald [287], the empirical contributions 1 and 3 are *knowledge-seeking* contributions, while the artifact and methodological

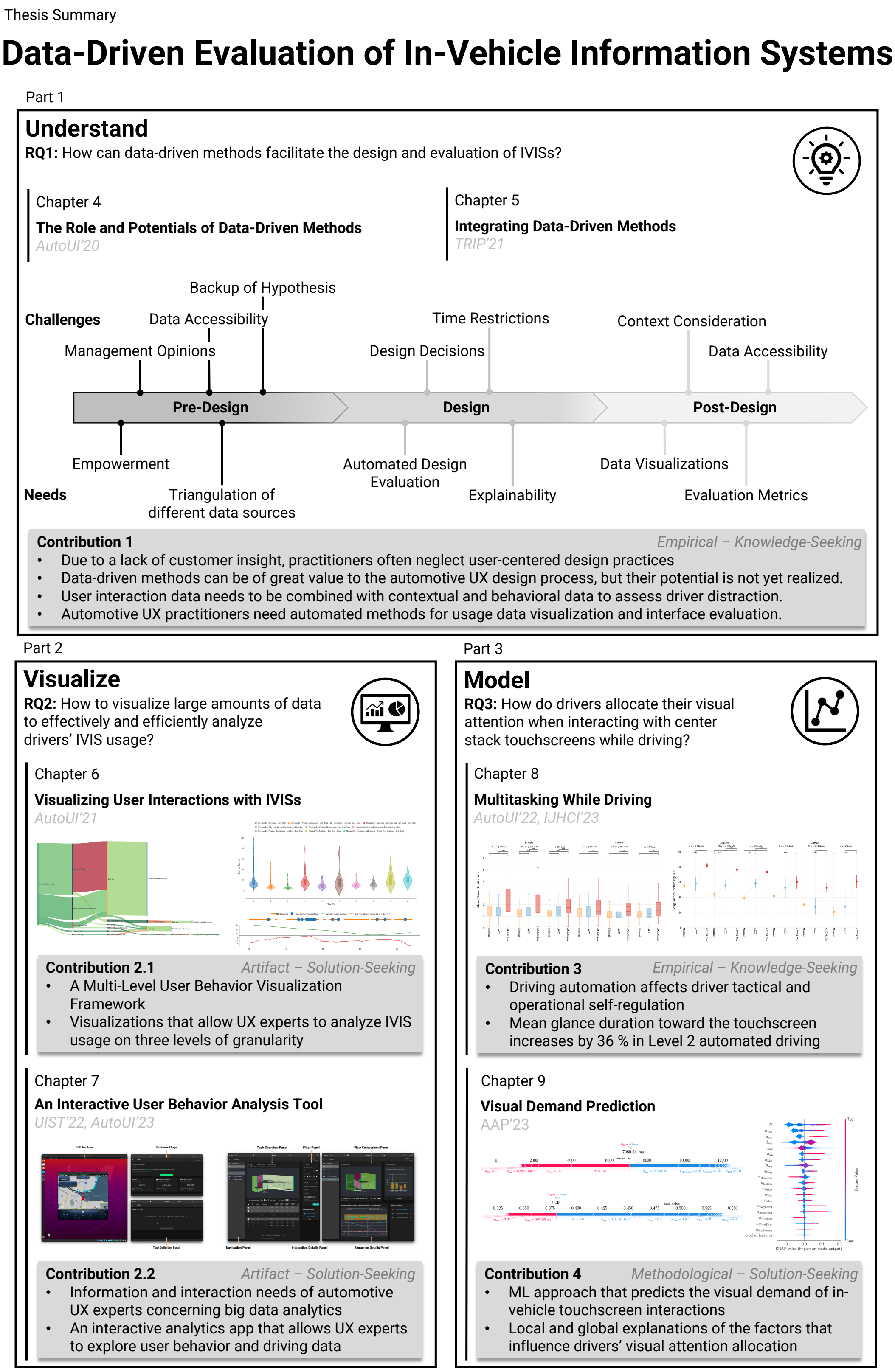


Figure 10.1: Visual summary highlighting the three main parts of this thesis.

contributions, 2 and 4, are *solution-seeking* contributions. While the knowledge seeking solutions inform the automotive UX design process in general and provide a substantial addition to the knowledge base, the solution-seeking contributions can be directly mapped to specific challenges and needs in the automotive UX design process. Contribution 2 focuses on the needs and challenges of the pre-design and post-design phases, while Contribution 3 focuses on the design phase. In the following, we answer the leading research questions and discuss how the individual contributions relate to each other.

RQ1: How Can Data-Driven Methods Facilitate the Design and Evaluation of IVISs?

The empirical research conducted in Part I shows that throughout the automotive design process, practitioners often neglect UCD practices and rely on subjective judgments instead of evidence-based user behavior insights. This is due to a lack of customer feedback. UX experts want to know how users interact with the various features the IVISs offer, how these interactions affect driving behavior and distraction, and how drivers adapt their behavior to the driving context. We found that the design and evaluation of IVISs is still mostly based on qualitative methods or small-scale user studies with human participants. However, these studies are costly and not scalable given limited resources and the complexity of the design task, leaving many questions unanswered. Furthermore, the results of these studies are often not valued and overruled by management decisions if they are not backed up by quantitative insights (see Figure 10.1). Therefore, traditional methods need to be complemented by data-driven methods that use large amounts of customer data to generate the anticipated customer insights. We argue that data-driven methods can shift the decision-making process away from personal intuition and best guesses to evidence-based design decisions, making the design and evaluation of IVISs more user-centered. However, UX experts face many challenges throughout the design process when it comes to the utilization of customer usage data (see Figure 10.1).

Interaction Data Needs to Be Combined with Contextual and Behavioral Data to Asses Driver Distraction While interacting with IVISs, drivers divide their attention between the primary driving task and the secondary IVIS task. Visually demanding secondary tasks, such as touchscreen interactions, are associated with an increased crash risk and drivers adjust their task-switching behavior according to the driving situation and the secondary task demands. Thus, IVIS evaluation becomes a safety-critical and context-sensitive task. Methods based on large amounts of data collected from customer vehicles must combine interaction data with driving data and glance data. This allows UX experts to assess driver distraction and provides a better understanding of how the driving demand affects drivers' glance and interaction behavior. Only then can IVISs be evaluated for their distraction potential in the context of the driving situation.

Visual Analytics to Accelerate Knowledge Generation and Improve Post-Release Testing Our results show that the usage data needed to make user-centered design decisions is often unavailable or poorly presented, preventing UX experts from using it to their advantage. This challenge especially applies to the knowledge generation processes in the pre-design phase and post-release testing in the post-design phase. Without access to properly prepared usage data, UX experts are often uncertain when it comes to understanding users, their tasks, and the context in which they use the IVIS. As a result, requirements elicitation is not truly user-centered. Without customer usage data, it is impossible to evaluate how drivers interact with the final product and whether it meets the requirements. Our research shows that visual analytics tools that combine multiple data sources and automate data processing and visualization can address these challenges. Tailored to the specific needs of automotive UX experts, they enable the exploration and evaluation of customer interactions with IVISs, empowering experts in early and late design phases.

Model-Based Evaluations Can Offer Valuable Design Support During the design phase, UX experts often lack the resources to evaluate their prototypes with real users. This leads to prototypes mainly being evaluated qualitatively by co-workers or in-house UX experts. While evaluations with experts can provide important insights, they are not suitable to objectively evaluate the usability or distraction potential of IVISs. However, feedback on metrics such as time on task or visual demand can be valuable in the early stages of the design process. Our results show that UX experts need model-based usability and distraction evaluations to inform design decisions. On the one hand, statistical models can generate knowledge about the complex interdependencies between driver interactions, driving context, and drivers' visual attention allocation. On the other hand, computational models that predict and simulate human behavior can be used to automatically evaluate early-stage prototypes.

RQ2: How to Visualize Large Amounts of Data to Effectively and Efficiently Analyze Drivers' IVIS Usage?

Our answer to this question, as presented in Part II, is twofold: First, we developed three visualizations according to the needs of the UX experts we identified in our interview study in Chapter 4. Second, we integrated these visualizations into a tool that automates data processing and visualization to enable UX experts to effectively and efficiently explore the data and analyze how drivers interact with the center stack touchscreen.

Create Automotive-Specific Visualizations In Chapter 6, we present a Multi-Level User Behavior Visualization Framework consisting of three visualizations that allow UX experts to explore drivers' center stack touchscreen interaction on different levels of detail. The visualizations serve the needs of automotive UX experts to visualize user interactions alongside glance and driving data. The three levels of detail allow UX experts to explore how users solve IVIS tasks, how different user flows compare according to performance and distraction metrics, and how drivers act in varying situations. Our evaluation shows that these visualizations benefit the design process of IVISs. They enable UX experts to explore large amounts of usage data and help them to find usability problems and unexpected user behavior.

Develop Interactive Tools That Automate Data Processing and Visualization Generation In Chapter 7, we present ICEBOAT, an interactive domain-specific visualization tool that allows automotive UX experts to analyze driver interactions with the center stack touchscreen. ICEBOAT automates the data processing and visualization generation so that UX experts have constant and immediate access to usage data that is continuously collected from production line vehicles via a Telematics Data Logging Framework. ICEBOAT enhances IVIS interaction data with driving and glance data and combines the visualizations developed in Chapter 6 such that they serve the needs of UX experts. Our tool supports the definition of user tasks and introduces several metrics and filters that allow UX experts to compare different flows, cars, or software versions according to usability factors and safety-related metrics. In our evaluation study, the tool proved to be intuitive and valuable in the context of the automotive UX design process. It enables UX experts to make data-driven decisions in situations where they would otherwise have to rely on their intuition.

RQ3: How Do Drivers Allocate Their Visual Attention When Interacting with Center Stack Touchscreens While Driving?

In Part III, we present two different approaches to answer the question of how drivers allocate their visual attention when interacting with center stack touchscreens. The first approach, presented in Chapter 8, is an empirical study. We apply statistical models to a naturalistic driving dataset to investigate drivers' tactical and operational self-regulation. While the first approach relies on historical data and contributes new findings to the field of driver distraction research, the second approach (Chapter 9) presents a method that can predict the visual demand of touchscreen interactions with early IVIS prototypes based on continuously collected customer data.

Drivers Self-Regulate Their Interactions with In-Vehicle Touchscreens on the Tactical and Operational Level In Chapter 8 we present the first naturalistic driving study to investigate drivers' tactical and operational self-regulation with center stack touchscreens. To understand how drivers adapt their engagement in secondary touchscreen tasks, we apply various mixed-effects models on a naturalistic driving dataset consisting of more than 30,000 secondary task engagements. Our results show significant differences in driver interaction and glance behavior in response to different levels of driving automation, vehicle speed, and road curvature. During level 2 automated driving, drivers perform more interactions per touchscreen sequence (+17 % compared to manual driving) and increase the time spent looking at the center stack touchscreen (+36 %). We also show that the effect of driving automation on the driver's self-regulation is greater than the effect of vehicle speed and road curvature. These findings have implications for the overall design process and highlight the importance of developing methods to evaluate early-stage prototypes for distraction potential.

Explainable Predictions for the Visual Demand of In-Vehicle Touchscreen Interactions The empirical study presented in Chapter 8 shows that drivers' visual attention allocation is influenced not only by the design of the touchscreen interface but also by the driving context. This demonstrates the importance of context-dependent distraction evaluation in the early stages of IVIS design. Thus, in Chapter 9, we present a machine learning-based approach to predict and explain the visual demand of in-vehicle touchscreen interactions based on large naturalistic driving data. By predicting the mean and total glance durations for yet unseen touchscreen interactions while driving, our method builds the basis for automated data-driven evaluations of early-stage IVIS prototypes. The local and global explanations generated by the SHAP method provide detailed insights into how design artifacts and driving context affect drivers' glance behavior. Our results are consistent with those from conventional user studies proving that our approach addresses the multitude of factors that influence in-vehicle interaction. Accordingly, our approach is a meaningful extension of traditional user research methods. On the one hand, designs can be evaluated early and continuously, and on the other hand, the insights can be used to conduct comparably expensive user studies in a targeted manner.

10.2 Limitations

This thesis successfully addresses several open challenges in the automotive UX design process and presents solutions that satisfy the needs for data-driven evaluation of IVISs formulated by UX experts. While our visualization and modeling approaches address our research questions and represent important contributions to academia and industry, several limitations remain that need to be considered. In the following, we discuss these remaining limitations and discuss possible improvements.

Data Availability All our approaches rely on the Telematics Data Logging Framework introduced in Section 3.1 and the availability of the required sensors in the car. However, as highlighted in Chapter 5, the technical infrastructure is often the limiting factor for OEMs to apply data-driven methods like the ones developed in conjunction with this thesis. Vehicles must be equipped with appropriate sensors, and automotive software platforms must be designed to support the collection of large amounts of data. However, this is not a trivial task. The long product life cycles in the automotive industry, the distribution of data across subsystems, and the current limitations of over-the-air updates pose significant challenges. In addition, privacy regulations can limit the use of data. These challenges must be overcome to enable widespread adoption of our approaches in the industry.

Anonymized Data From a Corporate Test Fleet Due to privacy regulations, the data used in this work is highly anonymized and is collected only from a company-owned test fleet. As a result, we cannot differentiate between individual drivers, nor do we have access to drivers' demographic information. We can only distinguish between different trips and car types. Considering that more than 100 cars with even more individual drivers contributed to the data collection, the risk of overfitting to individual drivers is low. However, because only company-owned test cars contributed to the data collection, the results are likely biased toward middle-aged drivers who are considered expert users. In addition, the software in these test cars is constantly evolving and frequently updated. This applies to the UI software as well as the camera or ADAS software. Changes in the software may affect the way drivers interact with the IVIS or how they self-regulate their behavior. These limitations are mainly relevant to the interpretation of the results presented in Chapter 8, but do not affect the artifact and methodological contributions.

Focus on Center Stack Touchscreens Our visualization and modeling approaches, as presented in Part II and Part III, are limited to touchscreen interactions with center stack IVISs. However, there are several other IVISs such as Head-Up Displays (HUDs), interactive instrument panels, and passenger displays that can influence driver behavior and distraction [288, 289]. Furthermore, drivers can interact with IVISs not only by touch, but also by voice, hard keys, or gestures. The mode of interaction affects user experience and driver distraction [290, 291, 292, 293]. We decided to focus on touchscreen interactions with the center stack IVIS as it contains the majority of the infotainment features available in today's cars. In addition, touchscreens are becoming more popular and are the primary interface between the driver and IVIS. However, as technology improves, this may change. Voice interactions already play an important role and gestures have found their way into the car. Therefore, it is important to go beyond the data-driven evaluation of center stack touchscreen interactions and consider additional systems and modalities.

Incomplete Descriptions of the Driving Context It is important to note that the features we use to describe the driving context are far from sufficient to capture all the factors that influence driving demand. For both modeling approaches, we consider the level of driving automation, vehicle speed, and steering wheel angle to describe the driving context. Our results show that these features provide valuable information. Yet, they do not allow for a comprehensive description. For example, the driving demand of driving at 100 km/h on a straight road may be significantly different in varying traffic situations or weather conditions. To evaluate the driving situation in more detail, additional information such as lighting conditions or traffic, weather or road type information may be valuable. The inclusion of additional features can not only improve the description of the driving context but also make the existing features more meaningful due to interaction effects. From an engineering perspective, adding new features to the models is straightforward. The data processing pipeline has to be adapted and the models have to be re-fitted. However, this can introduce new problems such as overfitting or loss of generalization due to the inclusion of irrelevant features [294]. Furthermore, a small number of meaningful features may be preferable as it increases the interpretability of the data [295].

10.3 Outlook

Our contributions open several future research directions and provide an entry point for various practical applications of data-driven methods to evaluate IVISs. In the following, we provide an outlook on how our approaches can be extended to better understand driver behavior and to design IVISs that are safe to use.

Personalized Predictions and Data-Driven Personas Research shows that driver characteristics such as age or health status and social-psychological factors such as attitudes, social norms, and personality affect drivers' self-regulation and distraction engagement [296, 297, 298]. Thus, being able to differentiate between individuals is likely to improve the fit and accuracy of the statistical and computational models. Furthermore, the ability to map driver characteristics (for example extracted from drivers' user profiles) with the data used in this work would open up various opportunities. One potential, as also identified in Section 5.2.3, is the ability to build archetypal user models that serve as data-driven personas [218, 219, 220]. As of now, personas are an abstract construct created using qualitative approaches [216] with no direct connection to real users' interaction or driving behavior [217]. Enhancing these personas with real-world interaction and driving data would allow to tailor and even adapt designs and functionalities to the characteristics and preferences of different user groups. Given the additional information, our approaches could be used to identify differences between user groups according to usage patterns, self-regulation, and visual attention allocation.

Visualization and Modeling of Multimodal Secondary Task Engagements As stated in Section 10.2, this work is currently limited to the modeling and visualization of touchscreen interactions. However, to holistically evaluate drivers' interaction behavior with IVISs it is important to not only evaluate touch interactions but to also evaluate interactions via voice, gestures, or haptic interfaces. The evaluation of multimodal interaction becomes particularly important as voice and gesture recognition improve and become a valid and potentially less distracting [292, 299] alternative. This is also reflected in the results presented in Chapter 4 and Chapter 7.

While early approaches exist that evaluate multimodal interactions with IVISs [300, 301], there are no data-driven approaches to visualize large amounts of multimodal interaction data or to model multimodal driver interactions. Since our approach offers the unique possibility to combine millions of voice and gesture commands from production vehicles with the data analyzed in this work, it would be a natural next step to extend our visualization and modeling approaches to multimodal interactions. Assuming that voice, gesture, and hard key interactions can be identified similarly to touchscreen interactions, from an engineering perspective it would be possible to adjust and extend the User Behavior Evaluation Module accordingly. However, despite the engineering effort, the evaluation of voice and gesture interactions raises several open questions. For example, gesture and voice interactions are not as directly coupled to IVIS functions as touchscreen interactions. These interactions need to be recognized, interpreted, and mapped to IVIS functions. This introduces several intermediate levels of information processing that can lead to errors and need to be evaluated accordingly. Furthermore, there has been little research on how to visualize multimodal interactions. While Jansen et al. [90] provide a first perspective on how individual sequences can be visualized to evaluate automotive user interfaces, it remains an open challenge how to aggregate and visualize large amounts of multimodal event sequences.

Abstract Representations of the Driving Context As shown in Chapter 9 and Chapter 8, drivers' interaction and glance behavior is highly sensitive to the driving context. Whereas our approaches partly represent this driving context, they are far from able to fully capture all factors that influence driving demand. To do so, more features need to be included. However, as pointed out in Section 10.2, adding more and more features can improve model performance but may reduce data interpretability. Thus, abstract descriptions of the driving scenario (e.g., *"rural area – straight – little traffic – good visibility"*) might prove more informative than simply adding more driving parameters. For this, one could take advantage of approaches to automatically understand driving scenes based on image segmentation and sensor fusion [302], such as those used in automated driving systems. Another approach to generating abstract representations of the driving context that can improve prediction accuracy could be provided by unsupervised feature learning [303].

Feature learning also referred to as representation learning is used to discover low-dimensional features that represent the underlying structure in high-dimensional data. These abstract descriptions of the driving context may not only improve prediction accuracy or model fit but may also be valuable for uncovering yet hidden dependencies between driver behavior and driving context.

Opening the Black Box To predict drivers' visual demand we rely on supervised machine learning (see Chapter 9). Although XAI approaches such as SHAP provide human-understandable explanations, they solely explain the output of the machine learning model. As there is no guarantee that the model itself represents human behavior, there is no guarantee that the explanations capture the real causal relationship between driver behavior and the driving environment. One way to address this challenge is to utilize the concept of computational rationality [304, 305]. Computationally rational theories are based on the assumption that users choose their behavior to maximize their expected utility, given their bounds [306]. Applied to human-computer interaction, Oulasvirta et al. [306] argue that users interact with technology so that they can achieve an optimal outcome given their internal (e.g., cognitive or perceptual) and external (e.g., environment or design of a tool) bounds. Thus, users adapt their behavioral policies or strategies to maximize their utility. These policies can be approximated via reinforcement learning, yielding verifiable predictions of user interaction behavior [306]. Although machine learning is used to learn behavioral policies and make predictions, the formulated bounds (e.g., perceptual or motor constraints) limit the space of computable interaction strategies of the agent so that the interactions represent realistic human behavior. This is the key difference to supervised machine learning models like those used in this work. While first microscopic approaches exist that utilize the concept of computational rationality to describe drivers' self-regulation [307, 308], much work remains to capture the broader facets of driver behavior such as modeling motivational aspects, interaction contexts, and situational awareness.

Bibliography

[1] P. Ebel, F. Brokhausen, and A. Vogelsang, "The Role and Potentials of Field User Interaction Data in the Automotive UX Development Lifecycle: An Industry Perspective," in *Proceedings of the 12th International Conference on Automotive User Interfaces and Interactive Vehicular Applications.* Virtual Event DC USA: ACM, Sep. 2020, pp. 141–150.

[2] P. Ebel, J. Orlovska, S. Hünemeyer, C. Wickman, A. Vogelsang, and R. Söderberg, "Automotive UX design and data-driven development: Narrowing the gap to support practitioners," *Transportation Research Interdisciplinary Perspectives*, vol. 11, p. 100455, Sep. 2021.

[3] P. Ebel, C. Lingenfelder, and A. Vogelsang, "Visualizing Event Sequence Data for User Behavior Evaluation of In-Vehicle Information Systems," in *Proceedings of the 13th International Conference on Automotive User Interfaces and Interactive Vehicular Applications.* Leeds United Kingdom: ACM, Sep. 2021, pp. 219–229.

[4] P. Ebel, M. Berger, C. Lingenfelder, and A. Vogelsang, "How Do Drivers Self-Regulate their Secondary Task Engagements? The Effect of Driving Automation on Touchscreen Interactions and Glance Behavior," in *Proceedings of the 14th International Conference on Automotive User Interfaces and Interactive Vehicular Applications.* Seoul Republic of Korea: ACM, Sep. 2022, pp. 263–273.

[5] P. Ebel, K. J. Gülle, C. Lingenfelder, and A. Vogelsang, "ICEBOAT: An Interactive User Behavior Analysis Tool for Automotive User Interfaces," in *The Adjunct Publication of the 35th Annual ACM Symposium on User Interface Software and Technology*, Aug. 2022.

[6] P. Ebel, C. Lingenfelder, and A. Vogelsang, "On the forces of driver distraction: Explainable predictions for the visual demand of in-vehicle touchscreen interactions," *Accident Analysis & Prevention*, vol. 183, p. 106956, Apr. 2023.

[7] P. Ebel, C. Lingenfelder, and A. Vogelsang, "Multitasking while driving: How drivers self-regulate their interaction with in-vehicle touchscreens in automated driving," *International Journal of Human–Computer Interaction*, pp. 1–18, 2023.

[8] P. Ebel, K. J. Gülle, C. Lingenfelder, and A. Vogelsang, "Exploring Millions of User Interactions with ICEBOAT: Big Data Analytics for Automotive User Interfaces," in *Proceedings of the 15th International Conference on Automotive User Interfaces and Interactive Vehicular Applications*, Ingolstadt, Germany, 2023.

[9] O. Fenzl, T. In der Smitten, G. Walthart, and C. Müller, "Mercedes-Benz Reveals New Charging Network and Tech Updates at CES 2023," https://group-media.mercedes-benz.com/marsMediaSite/ko/en/54924712, Jan. 2023.

[10] C. Harvey, N. A. Stanton, C. A. Pickering, M. McDonald, and P. Zheng, "In-Vehicle Information Systems to Meet the Needs of Drivers," *International Journal of Human-Computer Interaction*, vol. 27, no. 6, pp. 505–522, Jun. 2011.

[11] C. Harvey and N. A. Stanton, *Usability Evaluation for In-Vehicle Systems*, zeroth ed. CRC Press, Apr. 2016.

[12] G. Meixner, C. Häcker, B. Decker, S. Gerlach, A. Hess, K. Holl, A. Klaus, D. Lüddecke, D. Mauser, M. Orfgen, M. Poguntke, N. Walter, and R. Zhang, "Retrospective and Future Automotive Infotainment Systems—100 Years of User Interface Evolution," in *Automotive User Interfaces*, G. Meixner and C. Müller, Eds. Cham: Springer International Publishing, 2017, pp. 3–53.

[13] M. Kakihara and T. Asoh, "JAMA guideline for in-vehicle display systems (*Japan automobile manufacturers association)." Japan Automobile Manufacturers Association, 2004.

[14] National Center for Statistics and Analysis, "Distracted driving 2019," National Highway Traffic Safety Administration, Tech. Rep. Report No. DOT HS 811 299, 2021.

[15] N. H. T. S. Administration, "Visual-Manual NHTSA Driver Distraction Guidelines for In-Vehicle Electronic Devices," Sep. 2014.

[16] S. Klauer, T. Dingus, T. Neale, J. Sudweeks, and D. Ramsey, "The impact of driver inattention on near-crash/crash risk: An analysis using the 100-Car naturalistic driving study data," U.S. Department of Transportation, National Highway Traffic Safety Administration / Virginia Tech Transportation Institute, 3500 Transportation Research Plaza (0536) Blacksburg, Virginia 24061, Tech. Rep., Jan. 2006.

[17] J. Markoff, "Apple Introduces Innovative Cellphone," *The New York Times*, 2007.

[18] D. R. Large, G. Burnett, E. Crundall, G. Lawson, and L. Skrypchuk, "Twist It, Touch It, Push It, Swipe It: Evaluating Secondary Input Devices for Use with an Automotive Touchscreen HMI," in *Proceedings of the 8th International Conference on Automotive User Interfaces and Interactive Vehicular Applications*. Ann Arbor MI USA: ACM, Oct. 2016, pp. 161–168.

[19] "SAEJ3016: Taxonomy and definitions for terms related to driving automation systems for on-road motor vehicles," Society of Automotive Engineers (SAE), Warrendale, Standard, 2021.

[20] F. Faber, E. Jonkers, M. van Noort, M. Benmimoun, P. Andreas, B. Metz, G. Saint Pierre, D. Gustafson, and L. Malta, "EuroFOT deliverable 6.4 - final results: Impacts on traffic safety," ERTICO - ITS Europe, Project Report, 2012.

[21] R. Ervin, J. Sayer, D. LeBlanc, S. Bogard, M. Mefford, Z. Hagan, M. Bareket, and C. Winkler, "Automotive collision avoidance system field operational test report: Methodology and results," National Highway Traffic Safety Administration, Tech. Rep. HS-809 900, 2005.

[22] M. Risteska, D. Kanaan, B. Donmez, and H.-Y. W. Chen, "The effect of driving demands on distraction engagement and glance behaviors: Results from naturalistic data," *Safety Science*, vol. 136, p. 105123, Apr. 2021.

[23] A. Morando, T. Victor, and M. Dozza, "A Reference Model for Driver Attention in Automation: Glance Behavior Changes During Lateral and Longitudinal Assistance," *IEEE Transactions on Intelligent Transportation Systems*, vol. 20, no. 8, pp. 2999–3009, Aug. 2019.

[24] N. Dunn, T. Dingus, and S. Soccolich, "What's in a name? Drivers' perceptions of the use of five SAE Level 2 driving automation systems," *Journal of Safety Research*, vol. 72, pp. 145–151, Feb. 2020.

[25] M. Christoph, S. Wesseling, and N. van Nes, "Self-regulation of drivers' mobile phone use: The influence of driving context," *Transportation Research Part F: Traffic Psychology and Behaviour*, vol. 66, pp. 262–272, Oct. 2019.

[26] D. Onate-Vega, O. Oviedo-Trespalacios, and M. J. King, "How drivers adapt their behaviour to changes in task complexity: The role of secondary task demands and road environment factors," *Transportation Research Part F: Traffic Psychology and Behaviour*, vol. 71, pp. 145–156, May 2020.

[27] O. Oviedo-Trespalacios, M. M. Haque, M. King, and S. Washington, ""Mate! I'm running 10 min late": An investigation into the self-regulation of mobile phone tasks while driving," *Accident Analysis & Prevention*, vol. 122, pp. 134–142, Jan. 2019.

[28] C. A. DeGuzman and B. Donmez, "Knowledge of and trust in advanced driver assistance systems," *Accident Analysis & Prevention*, vol. 156, p. 106121, Jun. 2021.

[29] A. P. Hungund, G. Pai, and A. K. Pradhan, "Systematic Review of Research on Driver Distraction in the Context of Advanced Driver Assistance Systems," *Transportation Research Record: Journal of the Transportation Research Board*, vol. 2675, no. 9, pp. 756–765, Sep. 2021.

[30] On-Road Automated Driving (ORAD) committee, "Taxonomy and Definitions for Terms Related to Driving Automation Systems for On-Road Motor Vehicles," SAE International, Tech. Rep., 2021.

[31] S. Bryant and C. Wrigley, "Driving Toward User-Centered Engineering in Automotive Design," *Design Management Journal*, vol. 9, no. 1, pp. 74–84, Oct. 2014.

[32] J.-Y. Mao, K. Vredenburg, P. W. Smith, and T. Carey, "The state of user-centered design practice," *Communications of the ACM*, vol. 48, no. 3, pp. 105–109, Mar. 2005.

[33] R. Atterer, M. Wnuk, and A. Schmidt, "Knowing the user's every move: User activity tracking for website usability evaluation and implicit interaction," in *Proceedings of the 15th International Conference on World Wide Web - WWW '06*. Edinburgh, Scotland: ACM Press, 2006, p. 203.

[34] I. Pettersson, F. Lachner, A.-K. Frison, A. Riener, and A. Butz, "A Bermuda Triangle? A Review of Method Application and Triangulation in User Experience Evaluation," in *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, ser. CHI '18. New York, NY, USA: Association for Computing Machinery, 2018.

[35] J. Power, "2020 U.S. Initial Quality Study," J.D. Power, Tech. Rep., 2020.

[36] J. Power, "2021 U.S. Initial Quality Study," J.D. Power, Tech. Rep., 2021.

[37] J. Power, "2022 U.S. Initial Quality Study," J.D. Power, Tech. Rep., 2022.

[38] J. Orlovska, C. Wickman, and R. Soderberg, "The Use of Vehicle Data in ADAS Development, Verification and Follow-Up on the System," *Proceedings of the Design Society: DESIGN Conference*, vol. 1, pp. 2551–2560, May 2020.

[39] I. S. 39, "ISO 15007:2020 Road vehicles — Measurement and analysis of driver visual behaviour with respect to transport information and control systems," International Organization for Standardization, Geneva, CH, Standard, Mar. 2020.

[40] M. A. Regan, J. D. Lee, and K. L. Young, Eds., *Driver Distraction: Theory, Effects, and Mitigation.* Boca Raton: CRC Press/Taylor & Francis Group, 2009.

[41] E. L.-C. Law, V. Roto, M. Hassenzahl, A. P. O. S. Vermeeren, and J. Kort, "Understanding, scoping and defining user experience," in *Proceedings of the 27th International Conference on Human Factors in Computing Systems - CHI 09.* ACM Press, Apr. 2009, pp. 719–728.

[42] M. Hassenzahl and N. Tractinsky, "User experience - a research agenda," *Behaviour & Information Technology*, vol. 25, no. 2, pp. 91–97, Mar. 2006.

[43] J. Forlizzi and K. Battarbee, "Understanding experience in interactive systems," in *Proceedings of the 5th Conference on Designing Interactive Systems: Processes, Practices, Methods, and Techniques.* Cambridge MA USA: ACM, Aug. 2004, pp. 261–268.

[44] V. Roto, H. Rantavuo, and K. Väänänen, "Evaluating user experience of early product concepts," *Proceeding of the International Conference on Designing Pleasurable Products and Interfaces DPPI09*, 2009.

[45] K. Väänänen, V. Roto, and M. Hassenzahl, "Towards Practical User Experience Evaluation Methods," Jan. 2008.

[46] M. Hassenzahl, M. Burmester, and F. Koller, "User Experience Is All There Is: Twenty Years of Designing Positive Experiences and Meaningful Technology," *i-com*, vol. 20, no. 3, pp. 197–213, Dec. 2021.

[47] M. Hassenzahl, "User experience (UX): Towards an experiential perspective on product quality," in *Proceedings of the 20th Conference on l'Interaction Homme-Machine.* Metz France: ACM, Sep. 2008, pp. 11–15.

[48] j. Brooke, "SUS: A 'Quick and Dirty' Usability Scale," in *Usability Evaluation in Industry.* London, UK: Taylor & Francis, 1996, pp. 189–194.

[49] I. S. 4, "ISO 9241-11:2018 Ergonomics of human-system interaction — Part 11: Usability: Definitions and concepts," International Organization for Standardization, Tech. Rep., Mar. 2018.

[50] I. S. 4, "ISO 9241-210:2019 Ergonomics of human-system interaction — Part 210: Human-centred design for interactive systems," International Organization for Standardization, Tech. Rep., 2019.

[51] A. P. O. S. Vermeeren, E. L.-C. Law, V. Roto, M. Obrist, J. Hoonhout, and K. Väänänen-Vainio-Mattila, "User experience evaluation methods: Current state and development needs," in *Proceedings of the 6th Nordic Conference on Human-Computer Interaction: Extending Boundaries.* Reykjavik Iceland: ACM, Oct. 2010, pp. 521–530.

[52] E. L.-C. Law, "The measurability and predictability of user experience," in *Proceedings of the 3rd ACM SIGCHI Symposium on Engineering Interactive Computing Systems - EICS '11.* ACM Press, 2011.

[53] C. Harvey, N. A. Stanton, C. A. Pickering, M. McDonald, and P. Zheng, "Context of use as a factor in determining the usability of in-vehicle devices," *Theoretical Issues in Ergonomics Science*, vol. 12, no. 4, pp. 318–338, Jun. 2010.

[54] A. Löcken, S. S. Borojeni, H. Müller, T. M. Gable, S. Triberti, C. Diels, C. Glatz, I. Alvarez, L. Chuang, and S. Boll, "Towards Adaptive Ambient In-Vehicle Displays and Interactions: Insights and Design Guidelines from the 2015 AutomotiveUI Dedicated Workshop," in *Automotive User Interfaces.* Springer International Publishing, 2017, pp. 325–348.

[55] P. Fastrez and J.-B. Haué, "Designing and evaluating driver support systems with the user in mind," *International Journal of Human-Computer Studies*, vol. 66, no. 3, pp. 125–131, Mar. 2008.

[56] F. Chen and J. Terken, *Design Process.* Singapore: Springer Nature Singapore, 2023, pp. 165–179.

[57] J. Nielsen, "The usability engineering life cycle," *Computer*, vol. 25, no. 3, pp. 12–22, 1992.

[58] K. Vredenburg, J.-Y. Mao, P. W. Smith, and T. Carey, "A survey of user-centered design practice," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems.* Minneapolis Minnesota USA: ACM, Apr. 2002, pp. 471–478.

[59] X. Bi, A. Howes, P. O. Kristensson, A. Oulasvirta, and J. Williamson, *Introduction.* Oxford University Press, Mar. 2018, vol. 1.

[60] J. R. Saura, "Using Data Sciences in Digital Marketing: Framework, methods, and performance metrics," *Journal of Innovation & Knowledge*, vol. 6, no. 2, pp. 92–102, Apr. 2021.

[61] J. R. Saura, D. Ribeiro-Soriano, and D. Palacios-Marqués, “From user-generated data to data-driven innovation: A research agenda to understand user privacy in digital markets,” *International Journal of Information Management*, vol. 60, p. 102331, Oct. 2021.

[62] Y. Guo, S. Guo, Z. Jin, S. Kaul, D. Gotz, and N. Cao, “Survey on Visual Analysis of Event Sequence Data,” *IEEE Transactions on Visualization and Computer Graphics*, vol. 28, no. 12, pp. 5091–5112, Dec. 2022.

[63] D. I. Mattos, J. Bosch, and H. H. Olsson, “Your System Gets Better Every Day You Use It: Towards Automated Continuous Experimentation,” in *2017 43rd Euromicro Conference on Software Engineering and Advanced Applications (SEAA)*, Aug. 2017, pp. 256–265.

[64] G. Schermann, J. Cito, and P. Leitner, “Continuous Experimentation: Challenges, Implementation Techniques, and Current Research,” *IEEE Software*, vol. 35, no. 2, pp. 26–31, Mar. 2018.

[65] C. Dremel, J. Wulf, M. M. Herterich, J.-C. Waizmann, and W. Brenner, “How AUDI AG established big data analytics in its digital transformation.” *MIS Quarterly Executive*, vol. 16, no. 2, 2017.

[66] J. Orlovska, C. Wickman, and R. Söderberg, “Big Data Usage Can Be a Solution for User Behavior Evaluation: An Automotive Industry Example,” *Procedia CIRP*, vol. 72, pp. 117–122, 2018.

[67] F. Provost and T. Fawcett, “Data Science and its Relationship to Big Data and Data-Driven Decision Making,” *Big Data*, vol. 1, no. 1, pp. 51–59, Mar. 2013.

[68] E. Brynjolfsson, L. M. Hitt, and H. H. Kim, “Strength in Numbers: How Does Data-Driven Decisionmaking Affect Firm Performance?” *SSRN Electronic Journal*, 2011.

[69] J. W. Creswell, *Research Design: Qualitative, Quantitative, and Mixed Methods Approaches*. SAGE Publications, Mar. 2013.

[70] S. Merriam, *Qualitative Research : A Guide to Design and Implementation*. San Francisco: Jossey-Bass, 2009.

[71] N. K. Deniz and Y. S. Lincoln, *The SAGE Handbook of Qualitative Research*. Sage Publications Ltd., May 2017.

[72] M. Hennink, I. Hutter, and A. Bailey, *Qualitative Research Methods*, 2nd ed. Thousand Oaks: SAGE Publications Ltd, 2019.

[73] P. Wintersberger, A.-K. Frison, A. Riener, and T. von Sawitzky, “Fostering User Acceptance and Trust in Fully Automated Vehicles: Evaluating the Potential of Augmented Reality,” *Presence: Teleoperators and Virtual Environments*, vol. 27, no. 1, pp. 46–62, Feb. 2018.

[74] M. Walch, T. Sieber, P. Hock, M. Baumann, and M. Weber, “Towards Cooperative Driving: Involving the Driver in an Autonomous Vehicle’s Decision Making,” in

Proceedings of the 8th International Conference on Automotive User Interfaces and Interactive Vehicular Applications. Ann Arbor MI USA: ACM, Oct. 2016, pp. 261–268.

[75] M. Berger, D. Dey, A. Dandekar, B. Barati, R. Bernhaupt, and B. Pfleging, "Together in the Car: A Comparison of Five Concepts to Support Driver-Passenger Collaboration," in *Proceedings of the 14th International Conference on Automotive User Interfaces and Interactive Vehicular Applications.* Seoul Republic of Korea: ACM, Sep. 2022, pp. 183–194.

[76] N. Merat, A. H. Jamson, F. C. H. Lai, and O. Carsten, "Highly Automated Driving, Secondary Task Performance, and Driver State," *Human Factors: The Journal of the Human Factors and Ergonomics Society*, vol. 54, no. 5, pp. 762–771, Oct. 2012.

[77] L. Fridman, D. E. Brown, M. Glazer, W. Angell, S. Dodd, B. Jenik, J. Terwilliger, A. Patsekin, J. Kindelsberger, L. Ding, S. Seaman, A. Mehler, A. Sipperley, A. Pettinato, B. D. Seppelt, L. Angell, B. Mehler, and B. Reimer, "MIT Advanced Vehicle Technology Study: Large-Scale Naturalistic Driving Study of Driver Behavior and Interaction With Automation," *IEEE Access*, vol. 7, pp. 102 021–102 038, 2019.

[78] T. Dingus, S. Klauer, V. Lewis, A. Petersen, S. Lee, J. Sudweeks, M. Perez, J. Hankey, D. Ramsey, S. Gupta, C. Bucher, Z. Doerzaph, J. Jermeland, and R. Knipling, "The 100-Car Naturalistic Driving Study: Phase II – Results of the 100-Car Field Experiment," Jan. 2006.

[79] T. Victor, M. Dozza, J. Bärgman, C.-N. Boda, J. Engström, and G. Markkula, "Analysis of naturalistic driving study data: Safer glances, driver inattention, and crash risk," 2015.

[80] R. Eenink, Y. Barnard, M. Baumann, X. Augros, and F. Utesch, "UDRIVE: The European naturalistic driving study," in *Transport Research Arena (TRA) 5th Conference: Transport Solutions from Research to Deployment*, 2014.

[81] D. Fisher, R. DeLine, M. Czerwinski, and S. Drucker, "Interactions with big data analytics," *Interactions*, vol. 19, no. 3, pp. 50–59, May 2012.

[82] W. Cui, "Visual Analytics: A Comprehensive Overview," *IEEE Access*, vol. 7, pp. 81 555–81 573, 2019.

[83] D. A. Keim, F. Mansmann, J. Schneidewind, J. Thomas, and H. Ziegler, "Visual analytics: Scope and challenges," in *Visual Data Mining.* Springer, 2008, pp. 76–90.

[84] D. Keim, G. Andrienko, J.-D. Fekete, C. Görg, J. Kohlhammer, and G. Melançon, "Visual Analytics: Definition, Process, and Challenges," in *Information Visualization*, A. Kerren, J. T. Stasko, J.-D. Fekete, and C. North, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2008, vol. 4950, pp. 154–175.

[85] L. Zhang, A. Stoffel, M. Behrisch, S. Mittelstadt, T. Schreck, R. Pompl, S. Weber, H. Last, and D. Keim, "Visual analytics for the big data era — A comparative review of state-of-the-art commercial systems," in *2012 IEEE Conference on*

Visual Analytics Science and Technology (VAST). Seattle, WA, USA: IEEE, Oct. 2012, pp. 173–182.

[86] C. Arbesser, T. Mühlbacher, S. Komornyik, and H. Piringer, "Visual analytics for domain experts: Challenges and lessons learned," in *Proceedings of the Second International Symposium on Virtual Reality & Visual Computing*, V. K. T. Science and L. Technology CO., Eds. VR Kebao (Tiajin) Science and Technology CO.,Ltd, 2017, pp. 1–6.

[87] C. Minard, "Minard.png," https://commons.wikimedia.org/wiki/File:Minard.png, 1869.

[88] K. Wongsuphasawat and D. Gotz, "Exploring Flow, Factors, and Outcomes of Temporal Event Sequences with the Outflow Visualization," *IEEE Transactions on Visualization and Computer Graphics*, vol. 18, no. 12, pp. 2659–2668, Dec. 2012.

[89] M. Sedlmair, P. Isenberg, D. Baur, M. Mauerer, C. Pigorsch, and A. Butz, "Cardiogram: Visual analytics for automotive engineers," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems.* Vancouver BC Canada: ACM, May 2011, pp. 1727–1736.

[90] P. Jansen, J. Britten, A. Häusele, T. Segschneider, M. Colley, and E. Rukzio, "AutoVis: Enabling Mixed-Immersive Analysis of Automotive User Interface Interaction Studies," 2023.

[91] Z. Liu, Y. Wang, M. Dontcheva, M. Hoffman, S. Walker, and A. Wilson, "Patterns and Sequences: Interactive Exploration of Clickstreams to Understand Common Visitor Paths," *IEEE transactions on visualization and computer graphics*, vol. 23, no. 1, pp. 321–330, 2017.

[92] G. Wang, X. Zhang, S. Tang, H. Zheng, and B. Y. Zhao, "Unsupervised Clickstream Clustering for User Behavior Analysis," in *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, ser. CHI '16. New York, NY, USA: Association for Computing Machinery, 2016, pp. 225–236.

[93] J. Zhao, Z. Liu, M. Dontcheva, A. Hertzmann, and A. Wilson, "MatrixWave: Visual comparison of event sequence data," in *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems.* New York, NY, USA: Association for Computing Machinery, 2015, pp. 259–268.

[94] D. Gotz and H. Stavropoulos, "DecisionFlow: Visual Analytics for High-Dimensional Temporal Event Sequence Data," *IEEE Transactions on Visualization and Computer Graphics*, vol. 20, no. 12, pp. 1783–1792, Dec. 2014.

[95] S. Guo, K. Xu, R. Zhao, D. Gotz, H. Zha, and N. Cao, "EventThread: Visual Summarization and Stage Analysis of Event Sequence Data," *IEEE Transactions on Visualization and Computer Graphics*, vol. 24, no. 1, pp. 56–65, Jan. 2018.

[96] M. Friendly, "Visions and Re-Visions of Charles Joseph Minard," *Journal of Educational and Behavioral Statistics*, vol. 27, no. 1, pp. 31–51, Mar. 2002.

[97] P. Riehmann, M. Hanfler, and B. Froehlich, "Interactive Sankey diagrams," in *IEEE Symposium on Information Visualization, 2005. INFOVIS 2005.* IEEE, 2005.

[98] K. Wongsuphasawat, J. A. Guerra Gómez, C. Plaisant, T. D. Wang, M. Taieb-Maimon, and B. Shneiderman, "LifeFlow: Visualizing an overview of event sequences," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems.* Vancouver BC Canada: ACM, May 2011, pp. 1747–1756.

[99] B. Deka, Z. Huang, C. Franzen, J. Nichols, Y. Li, and R. Kumar, "ZIPT: Zero-Integration Performance Testing of Mobile App Designs," in *Proceedings of the 30th Annual ACM Symposium on User Interface Software and Technology - UIST '17.* ACM Press, 2017.

[100] J. J. Thomas and K. A. Cook, "A visual analytics agenda," *IEEE computer graphics and applications*, vol. 26, no. 1, pp. 10–13, 2006.

[101] D. A. Keim, F. Mansmann, and J. Thomas, "Visual analytics: How much visualization and how much analytics?" *ACM SIGKDD Explorations Newsletter*, vol. 11, no. 2, pp. 5–8, May 2010.

[102] G. Shmueli, "To Explain or to Predict?" *Statistical Science*, vol. 25, no. 3, Aug. 2010.

[103] A. Oulasvirta and K. Hornbæk, "HCI Research as Problem-Solving," in *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems.* San Jose California USA: ACM, May 2016, pp. 4956–4967.

[104] C. P. Janssen, L. N. Boyle, W. Ju, A. Riener, and I. Alvarez, "Agents, environments, scenarios: A framework for examining models and simulations of human-vehicle interaction," *Transportation Research Interdisciplinary Perspectives*, vol. 8, p. 100214, Nov. 2020.

[105] I. S. MacKenzie, *Human-Computer Interaction: An Empirical Research Perspective*, 1st ed. Amsterdam: Morgan Kaufmann is an imprint of Elsevier, 2013.

[106] R. A. Johnson, I. Miller, and J. E. Freund, *Miller & Freund's Probability and Statistics for Engineers*, 9th ed. Boston columbus Indianapolis: Pearson, 2018.

[107] H. Grahn and T. Kujala, "Impacts of Touch Screen Size, User Interface Design, and Subtask Boundaries on In-Car Task's Visual Demand and Driver Distraction," *International Journal of Human-Computer Studies*, vol. 142, p. 102467, Oct. 2020.

[108] J. Hox, "Multilevel Modeling: When and Why," in *Classification, Data Analysis, and Data Highways*, H. H. Bock, O. Opitz, M. Schader, I. Balderjahn, R. Mathar, and M. Schader, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 1998, pp. 147–154.

[109] R. C. Wilson and A. G. Collins, "Ten simple rules for the computational modeling of behavioral data," *eLife*, vol. 8, p. e49547, Nov. 2019.

[110] N. Banovic, A. Oulasvirta, and P. O. Kristensson, "Computational Modeling in Human-Computer Interaction," in *Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems.* Glasgow Scotland Uk: ACM, May 2019, pp. 1–7.

[111] L. Breiman, "Statistical Modeling: The Two Cultures (with comments and a rejoinder by the author)," *Statistical Science*, vol. 16, no. 3, Aug. 2001.

[112] B. Deka, Z. Huang, C. Franzen, J. Hibschman, D. Afergan, Y. Li, J. Nichols, and R. Kumar, "Rico: A Mobile App Dataset for Building Data-Driven Design Applications," in *Proceedings of the 30th Annual ACM Symposium on User Interface Software and Technology*. ACM, Oct. 2017.

[113] K. Z. Gajos, D. S. Weld, and J. O. Wobbrock, "Automatically generating personalized user interfaces with Supple," *Artificial Intelligence*, vol. 174, no. 12-13, pp. 910–950, 2010.

[114] K. Todi, D. Weir, and A. Oulasvirta, "Sketchplore," in *Proceedings of the 2016 ACM Conference on Designing Interactive Systems*. ACM, Jun. 2016.

[115] A. Oulasvirta, N. R. Dayama, M. Shiripour, M. John, and A. Karrenbauer, "Combinatorial Optimization of Graphical User Interface Designs," *Proceedings of the IEEE*, vol. 108, no. 3, pp. 434–464, Mar. 2020.

[116] P. Ebel, I. E. Gol, C. Lingenfelder, and A. Vogelsang, "Destination Prediction Based on Partial Trajectory Data," in *2020 IEEE Intelligent Vehicles Symposium (IV)*. Las Vegas, NV, USA: IEEE, Oct. 2020, pp. 1149–1155.

[117] M. Baratchi, N. Meratnia, P. J. M. Havinga, A. K. Skidmore, and B. A. K. G. Toxopeus, "A hierarchical hidden semi-Markov model for modeling mobility data," in *Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing*. Seattle Washington: ACM, Sep. 2014, pp. 401–412.

[118] E. Pakdamanian, S. Sheng, S. Baee, S. Heo, S. Kraus, and L. Feng, "DeepTake: Prediction of Driver Takeover Behavior using Multimodal Data," in *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. Yokohama Japan: ACM, May 2021, pp. 1–14.

[119] D. Kanaan, S. Ayas, B. Donmez, M. Risteska, and J. Chakraborty, "Using Naturalistic Vehicle-Based Data to Predict Distraction and Environmental Demand," *International Journal of Mobile Human Computer Interaction*, vol. 11, no. 3, pp. 59–70, Jul. 2019.

[120] Z. Li, S. Bao, I. V. Kolmanovsky, and X. Yin, "Visual-Manual Distraction Detection Using Driving Performance Indicators With Naturalistic Driving Data," *IEEE Transactions on Intelligent Transportation Systems*, vol. 19, no. 8, pp. 2528–2535, Aug. 2018.

[121] M. A. Regan and O. Oviedo-Trespalacios, "Driver Distraction: Mechanisms, Evidence, Prevention, and Mitigation," in *The Vision Zero Handbook*, K. Edvardsson Björnberg, M.-Å. Belin, S. O. Hansson, and C. Tingvall, Eds. Cham: Springer International Publishing, 2022, pp. 1–62.

[122] J. Lee, K. Young, and M. Regan, "Defining Driver Distraction," in *Driver Distraction*, M. Regan, J. Lee, and K. Young, Eds. CRC Press, Oct. 2008, pp. 31–40.

[123] K. Kircher and C. Ahlstrom, "Minimum Required Attention: A Human-Centered Approach to Driver Inattention," *Human Factors: The Journal of the Human Factors and Ergonomics Society*, vol. 59, no. 3, pp. 471–484, May 2017.

[124] M. A. Regan, C. Hallett, and C. P. Gordon, "Driver distraction and driver inattention: Definition, relationship and taxonomy," *Accident Analysis & Prevention*, vol. 43, no. 5, pp. 1771–1781, Sep. 2011.

[125] World Health Organization and NHTSA (U.S.), "Mobile phone use: a growing problem of driver distraction," p. 48, 2011.

[126] T. A. Dingus, F. Guo, S. Lee, J. F. Antin, M. Perez, M. Buchanan-King, and J. Hankey, "Driver crash risk factors and prevalence evaluation using naturalistic driving data," *Proceedings of the National Academy of Sciences*, vol. 113, no. 10, pp. 2636–2641, Feb. 2016.

[127] S. G. Klauer, F. Guo, B. G. Simons-Morton, M. C. Ouimet, S. E. Lee, and T. A. Dingus, "Distracted Driving and Risk of Road Crashes among Novice and Experienced Drivers," *New England Journal of Medicine*, vol. 370, no. 1, pp. 54–59, Jan. 2014.

[128] G. M. Fitch, S. A. Soccolich, F. Guo, J. McClafferty, Y. Fang, R. L. Olson, M. A. Perez, R. J. Hanowski, J. M. Hankey, and T. A. Dingus, "The impact of hand-held and hands-free cell phone use on driving performance and safety-critical event risk," Tech. Rep., 2013.

[129] P. Gershon, K. R. Sita, C. Zhu, J. P. Ehsani, S. G. Klauer, T. A. Dingus, and B. G. Simons-Morton, "Distracted Driving, Visual Inattention, and Crash Risk Among Teenage Drivers," *American Journal of Preventive Medicine*, vol. 56, no. 4, pp. 494–500, Apr. 2019.

[130] M. Dozza, "What factors influence drivers' response time for evasive maneuvers in real traffic?" *Accident Analysis & Prevention*, vol. 58, pp. 299–308, Sep. 2013.

[131] P. Wintersberger, C. Schartmüller, S. Shadeghian-Borojeni, A.-K. Frison, and A. Riener, "Evaluation of imminent take-over requests with real automation on a test track," *Human Factors: The Journal of the Human Factors and Ergonomics Society*, p. 001872082110514, Dec. 2021.

[132] G. Merlhiot and M. Bueno, "How drowsiness and distraction can interfere with take-over performance: A systematic and meta-analysis review," *Accident Analysis & Prevention*, p. 106536, Dec. 2021.

[133] W. J. Horrey and C. D. Wickens, "In-vehicle glance duration: Distributions, tails and a model of crash risk," *Transportation Research Record: Journal of the Transportation Research Board*, vol. 2018, no. 1, pp. 22–28, Jan. 2007.

[134] C. Rudin-Brown, Ed., *Behavioural Adaptation and Road Safety: Theory, Evidence, and Action.* Boca Raton: CRC Press, Taylor & Francis Group, 2013.

[135] J. A. Michon, "A Critical View of Driver Behavior Models: What Do We Know, What Should We Do?" in *Human Behavior and Traffic Safety*, L. Evans and R. C. Schwing, Eds. Boston, MA: Springer US, 1985, pp. 485–524.

[136] K. L. Young and M. G. Lenné, "Driver engagement in distracting activities and the strategies used to minimise risk," *Safety Science*, vol. 48, no. 3, pp. 326–332, Mar. 2010.

[137] E. Tivesten and M. Dozza, "Driving context influences drivers' decision to engage in visual–manual phone tasks: Evidence from a naturalistic driving study," *Journal of Safety Research*, vol. 53, pp. 87–96, Jun. 2015.

[138] G. Hancox, J. Richardson, and A. Morris, "Drivers' willingness to engage with their mobile phone: The influence of phone function and road demand," *IET Intelligent Transport Systems*, vol. 7, no. 2, pp. 215–222, Jun. 2013.

[139] O. Oviedo-Trespalacios, M. M. Haque, M. King, and S. Washington, "Should I Text or Call Here? A Situation-Based Analysis of Drivers' Perceived Likelihood of Engaging in Mobile Phone Multitasking: Mobile Phone Multitasking Engagement," *Risk Analysis*, vol. 38, no. 10, pp. 2144–2160, Oct. 2018.

[140] R. Ismaeel, D. Hibberd, O. Carsten, and S. Jamson, "Do drivers self-regulate their engagement in secondary tasks at intersections? An examination based on naturalistic driving data," *Accident Analysis & Prevention*, vol. 137, p. 105464, Mar. 2020.

[141] W. J. Horrey and M. F. Lesch, "Driver-initiated distractions: Examining strategic adaptation for in-vehicle task initiation," *Accident Analysis & Prevention*, vol. 41, no. 1, pp. 115–122, Jan. 2009.

[142] O. Carsten, D. Hibberd, J. Bärgman, J. Kovaceva, M. S. Pereira Cocron, M. Dotzauer, F. Utesch, M. Zhang, E. Stemmler, L. Guyonvarch, F. Sagberg, and F. Forcolin, *UDRIVE Deliverable 43.1, Driver Distraction and Inattention, of the EU FP7 Project UDRIVE*, 1st ed. BE: UDRIVE Consortium, Jul. 2017.

[143] Y. Liang, W. J. Horrey, and J. D. Hoffman, "Reading Text While Driving: Understanding Drivers' Strategic and Tactical Adaptation to Distraction," *Human Factors: The Journal of the Human Factors and Ergonomics Society*, vol. 57, no. 2, pp. 347–359, Mar. 2015.

[144] T. Schneidereit, T. Petzoldt, A. Keinath, and J. F. Krems, "Using SHRP 2 naturalistic driving data to assess drivers' speed choice while being engaged in different secondary tasks," *Journal of Safety Research*, vol. 62, pp. 33–42, Sep. 2017.

[145] T. Morgenstern, L. Schott, and J. F. Krems, "Do drivers reduce their speed when texting on highways? A replication study using European naturalistic driving data," *Safety Science*, vol. 128, p. 104740, Aug. 2020.

[146] O. Oviedo-Trespalacios, M. M. Haque, M. King, and S. Demmel, "Driving behaviour while self-regulating mobile phone interactions: A human-machine system approach," *Accident Analysis & Prevention*, vol. 118, pp. 253–262, Sep. 2018.

[147] P. Choudhary and N. R. Velaga, "Mobile phone use during driving: Effects on speed and effectiveness of driver compensatory behaviour," *Accident Analysis & Prevention*, vol. 106, pp. 370–378, Sep. 2017.

[148] E. Tivesten and M. Dozza, "Driving context and visual-manual phone tasks influence glance behavior in naturalistic driving," *Transportation Research Part F: Traffic Psychology and Behaviour*, vol. 26, pp. 258–272, Sep. 2014.

[149] R. Lin, N. Liu, L. Ma, T. Zhang, and W. Zhang, "Exploring the self-regulation of secondary task engagement in the context of partially automated driving: A pilot study," *Transportation Research Part F: Traffic Psychology and Behaviour*, vol. 64, pp. 147–160, Jul. 2019.

[150] N. Schömig and B. Metz, "Three levels of situation awareness in driving with secondary tasks," *Safety Science*, vol. 56, pp. 44–51, Jul. 2013.

[151] J. Gaspar and C. Carney, "The Effect of Partial Automation on Driver Attention: A Naturalistic Driving Study," *Human Factors: The Journal of the Human Factors and Ergonomics Society*, vol. 61, no. 8, pp. 1261–1276, Dec. 2019.

[152] S. Yang, J. Kuo, and M. G. Lenné, "Effects of Distraction in On-Road Level 2 Automated Driving: Impacts on Glance Behavior and Takeover Performance," *Human Factors: The Journal of the Human Factors and Ergonomics Society*, vol. 63, no. 8, pp. 1485–1497, Dec. 2021.

[153] A. M. Noble, M. Miles, M. A. Perez, F. Guo, and S. G. Klauer, "Evaluating driver eye glance behavior and secondary task engagement while using driving automation systems," *Accident Analysis & Prevention*, vol. 151, p. 105959, Mar. 2021.

[154] A. Morando, P. Gershon, B. Mehler, and B. Reimer, "Visual attention and steering wheel control: From engagement to disengagement of Tesla Autopilot," *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, vol. 65, no. 1, pp. 1390–1394, Sep. 2021.

[155] J. Engström, E. Johansson, and J. Östlund, "Effects of visual and cognitive load in real and simulated motorway driving," *Transportation Research Part F: Traffic Psychology and Behaviour*, vol. 8, no. 2, pp. 97–120, Mar. 2005.

[156] B. Donmez, L. N. Boyle, and J. D. Lee, "Differences in off-road glances: Effects on young drivers' performance," *Journal of transportation engineering*, vol. 136, no. 5, pp. 403–409, 2010.

[157] Y. Liang and J. D. Lee, "Combining cognitive and visual distraction: Less than the sum of its parts," *Accident Analysis & Prevention*, vol. 42, no. 3, pp. 881–890, May 2010.

[158] P. Green, "Visual and task demands of driver information systems," The University of Michigan Transportation Research Institute (UMTRI), Tech. Rep., 1999.

[159] P. Burns, J. Harbluk, J. P. Foley, and L. Angell, "The importance of task duration and related measures in assessing the distraction potential of in-vehicle tasks," in *Proceedings of the 2nd International Conference on Automotive User Interfaces and Interactive Vehicular Applications - AutomotiveUI 10*. ACM Press, 2010.

[160] M. Wollmer, C. Blaschke, T. Schindl, B. Schuller, B. Farber, S. Mayer, and B. Trefflich, "Online driver distraction detection using long short-term memory," *IEEE Transactions on Intelligent Transportation Systems*, vol. 12, no. 2, pp. 574–582, Jun. 2011.

[161] L. Li, B. Zhong, C. Hutmacher, Y. Liang, W. J. Horrey, and X. Xu, "Detection of driver manual distraction via image-based hand and ear recognition," *Accident Analysis & Prevention*, vol. 137, p. 105432, Mar. 2020.

[162] M. Kutila, M. Jokela, G. Markkula, and M. R. Rue, "Driver distraction detection with a camera vision system," in *2007 IEEE International Conference on Image Processing*. IEEE, 2007.

[163] S. K. Card, T. P. Moran, and A. Newell, "The keystroke-level model for user performance time with interactive systems," *Communications of the ACM*, vol. 23, no. 7, pp. 396–410, Jul. 1980.

[164] S. Card, *The Psychology of Human-Computer Interaction.* Hillsdale, N.J: L. Erlbaum Associates, 1983.

[165] S. Schneegass, B. Pfleging, D. Kern, and A. Schmidt, "Support for modeling interaction with automotive user interfaces," in *Proceedings of the 3rd International Conference on Automotive User Interfaces and Interactive Vehicular Applications - AutomotiveUI '11.* ACM Press, 2011.

[166] D. I. Manes, "Prediction of destination entry and retrieval times using keystroke-level models," 1997.

[167] S. C. Lee, S. H. Yoon, and Y. G. Ji, "Modeling task completion time of in-vehicle information systems while driving with keystroke level modeling," *International Journal of Industrial Ergonomics*, vol. 72, pp. 252–260, Jul. 2019.

[168] M. Pettitt, G. Burnett, and A. Stevens, "An extended keystroke level model (KLM) for predicting the visual demand of in-vehicle information systems," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems.* ACM, Apr. 2007.

[169] J. W. Senders, AB. Kristofferson, WH. Levison, CW. Dietrich, JL. Ward *et al.*, "The attentional demand of automobile driving," 1967.

[170] C. Purucker, F. Naujoks, A. Prill, and A. Neukum, "Evaluating distraction of in-vehicle information systems while driving by predicting total eyes-off-road times with keystroke level modeling," *Applied Ergonomics*, vol. 58, pp. 543–554, Jan. 2017.

[171] D. R. Large, G. Burnett, E. Crundall, E. van Loon, A. L. Eren, and L. Skrypchuk, "Developing Predictive Equations to Model the Visual Demand of In-Vehicle Touchscreen HMIs," *International Journal of Human–Computer Interaction*, vol. 34, no. 1, pp. 1–14, Apr. 2017.

[172] D. Large, V. Banks, G. Burnett, S. Baverstock, and L. Skrypchuk, "Exploring the behaviour of distracted drivers during different levels of automation in driving,"

in *5th International Conference on Driver Distraction and Inattention (DDI2017)*, 2017.

[173] S. M. Pampel, G. Burnett, C. Hare, H. Singh, A. Shabani, L. Skrypchuk, and A. Mouzakitis, "Fitts goes autobahn: Assessing the visual demand of finger-touch pointing tasks in an on-road study," in *Proceedings of the 11th International Conference on Automotive User Interfaces and Interactive Vehicular Applications*, ser. AutomotiveUI '19. New York, NY, USA: Association for Computing Machinery, 2019, pp. 254–261.

[174] O. Tsimhoni and P. Green, "Visual Demand of Driving and the Execution of Display-Intensive in-Vehicle Tasks," *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, vol. 45, no. 23, pp. 1586–1590, Oct. 2001.

[175] T. Kujala and D. D. Salvucci, "Modeling visual sampling on in-car displays: The challenge of predicting safety-critical lapses of control," *International Journal of Human-Computer Studies*, vol. 79, pp. 66–78, Jul. 2015.

[176] J. R. Anderson, D. Bothell, M. D. Byrne, S. Douglass, C. Lebiere, and Y. Qin, "An Integrated Theory of the Mind." *Psychological Review*, vol. 111, no. 4, pp. 1036–1060, 2004.

[177] W. W. Wierwille, "Demands on driver resources associated with introducing advanced technology into the vehicle," *Transportation Research Part C: Emerging Technologies*, vol. 1, no. 2, pp. 133–142, Jun. 1993.

[178] K. Vassakis, E. Petrakis, and I. Kopanakis, "Big Data Analytics: Applications, Prospects and Challenges," in *Mobile Big Data*, G. Skourletopoulos, G. Mastorakis, C. X. Mavromoustakis, C. Dobre, and E. Pallis, Eds. Cham: Springer International Publishing, 2018, vol. 10, pp. 3–20.

[179] E. Brynjolfsson and K. McElheran, "The Rapid Adoption of Data-Driven Decision-Making," *American Economic Review*, vol. 106, no. 5, pp. 133–139, May 2016.

[180] LS. Angell, M. Perez, and J. Hankey, "Driver usage patterns for secondary information systems," in *Invited Paper for the First Human Factors Symposium on Naturalistic Driving Methods & Analyses*, 2008.

[181] J. Merchant, "The oculometer," National Aeronautics and Space Administration, Tech. Rep. NASA-CR-805, 1967.

[182] T. Hutchinson, K. White, W. Martin, K. Reichert, and L. Frey, "Human-computer interaction using eye-gaze input," *IEEE Transactions on Systems, Man, and Cybernetics*, vol. 19, no. 6, pp. 1527–1534, Nov.-Dec./1989.

[183] I. Etikan, "Comparison of convenience sampling and purposive sampling," *American journal of theoretical and applied statistics*, vol. 5, no. 1, pp. 1–4, 2016.

[184] J. Saldaña, *The Coding Manual for Qualitative Researchers*, 2nd ed. Los Angeles: SAGE, 2013.

[185] K. Krippendorff, *Content Analysis: An Introduction to Its Methodology.* Sage publications, 2018.

[186] G. Guest, A. Bunce, and L. Johnson, "How Many Interviews Are Enough?: An Experiment with Data Saturation and Variability," *Field Methods*, vol. 18, no. 1, pp. 59–82, Feb. 2006.

[187] J. Maxwell, *Qualitative Research Design: An Interactive Approach.* SAGE (London), 2012.

[188] J. Lewis, "Redefining Qualitative Methods: Believability in the Fifth Moment," *International Journal of Qualitative Methods*, vol. 8, no. 2, pp. 1–14, 2009.

[189] F. Benevenuto, T. Rodrigues, M. Cha, and V. Almeida, "Characterizing user behavior in online social networks," in *Proceedings of the 9th ACM SIGCOMM Conference on Internet Measurement Conference*, A. Feldmann, Ed. New York, NY: ACM, 2009, p. 49.

[190] M. Speicher, A. Both, and M. Gaedke, "Inuit: The Interface Usability Instrument," in *Design, User Experience, and Usability: Design Discourse.* Springer International Publishing, 2015, pp. 256–268.

[191] X. Ma, B. Yan, G. Chen, C. Zhang, K. Huang, J. Drury, and L. Wang, "Design and Implementation of a Toolkit for Usability Testing of Mobile Apps," *Mobile Networks and Applications*, vol. 18, no. 1, pp. 81–97, Nov. 2012.

[192] E. de Salis, D. Y. Baumgartner, and S. Carrino, "Can we predict driver distraction without driver psychophysiological state?" in *Proceedings of the 11th International Conference on Automotive User Interfaces and Interactive Vehicular Applications Adjunct Proceedings - AutomotiveUI '19.* ACM Press, 2019.

[193] M. Risteska, J. Chakraborty, and B. Donmez, "Predicting Environmental Demand and Secondary Task Engagement using Vehicle Kinematics from Naturalistic Driving Data," in *Proceedings of the 10th International Conference on Automotive User Interfaces and Interactive Vehicular Applications - AutomotiveUI '18.* ACM Press, 2018.

[194] C. Zhang, M. Patel, S. Buthpitiya, K. Lyons, B. Harrison, and G. D. Abowd, "Driver Classification Based on Driving Behaviors," in *Proceedings of the 21st International Conference on Intelligent User Interfaces - IUI '16.* ACM Press, 2016.

[195] J. Orlovska, C. Wickman, and R. Söderberg, "Big Data Analysis as a new Approach for Usability Attributes Evaluation of User Interfaces An Automotive Industry Context," in *Proceedings of the DESIGN 2018 15th International Design Conference*, ser. Design Conference Proceedings. Faculty of Mechanical Engineering and Naval Architecture, University of Zagreb, Croatia and The Design Society, Glasgow, UK, 2018, pp. 1651–1662.

[196] L. Lamm and C. Wolff, "Exploratory Analysis of the Research Literature on Evaluation of In-Vehicle Systems Interfaces and Interactive Vehicular Applications, AutomotiveUI 2019, Utrecht, The Netherlands, September 21-25, 2019," in *Proceedings of the 11th International Conference on Automotive User Interfaces and Interactive Vehicular Applications, AutomotiveUI 2019, Utrecht, The Netherlands, September 21-25, 2019*, C. P. Janssen, S. F. Donker, L. L. Chuang, and W. Ju, Eds. ACM, 2019, pp. 60–69.

[197] D. E. Avison, F. Lau, M. D. Myers, and P. A. Nielsen, "Action research," *Communications of the ACM*, vol. 42, no. 1, pp. 94–97, 1999.

[198] D. I. K. Sjoberg, T. Dyba, and M. Jorgensen, "The Future of Empirical Methods in Software Engineering Research," in *Future of Software Engineering (FOSE '07)*. IEEE, May 2007.

[199] D. Greenwood, *Introduction to Action Research : Social Research for Social Change*. Thousand Oaks: Sage Publications, 1998.

[200] J. Orlovska, F. Novakazi, B. Lars-Ola, M. Karlsson, C. Wickman, and R. Söderberg, "Effects of the driving context on the usage of Automated Driver Assistance Systems (ADAS) -Naturalistic Driving Study for ADAS evaluation," *Transportation Research Interdisciplinary Perspectives*, p. 100093, Feb. 2020.

[201] K. Petersen, C. Gencel, N. Asghari, D. Baca, and S. Betz, "Action research as a model for industry-academia collaboration in the software engineering context," in *Proceedings of the 2014 International Workshop on Long-term Industrial Collaboration on Software Engineering*. ACM, Sep. 2014.

[202] D. Collier and J. Mahoney, "Insights and Pitfalls: Selection Bias in Qualitative Research," *World Politics*, vol. 49, no. 1, pp. 56–91, 1996.

[203] M. Broy, "Challenges in Automotive Software Engineering," in *Proceedings of the 28th International Conference on Software Engineering*, ser. ICSE '06. New York, NY, USA: Association for Computing Machinery, 2006, pp. 33–42.

[204] A. Vogelsang, "Feature dependencies in automotive software systems: Extent, awareness, and refactoring," *Journal of Systems and Software (JSS)*, vol. 160, 2020.

[205] W. G. Voss and K. Houser, "Personal Data and the GDPR: Providing a Competitive Advantage for U.S. Companies," *American Business Law Journal*, vol. 56, pp. 287–344, 2019.

[206] E. Pernot-Leplay, "China's Approach on Data Privacy Law: A Third Way Between the US and the EU?" *Penn State Journal of Law & International Affairs*, vol. 8, no. 1, 2020.

[207] R. Kohavi, A. Deng, B. Frasca, T. Walker, Y. Xu, and N. Pohlmann, "Online controlled experiments at large scale," in *Proceedings of the 19th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. ACM, Aug. 2013.

[208] Y. Xu, N. Chen, A. Fernandez, O. Sinno, and A. Bhasin, “From Infrastructure to Culture,” in *Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining.* ACM, Aug. 2015.

[209] F. Giaimo, H. Andrade, and C. Berger, “Continuous experimentation and the cyber–physical systems challenge: An overview of the literature and the industrial perspective,” *Journal of Systems and Software*, vol. 170, p. 110781, Dec. 2020.

[210] I. Tesla, “Software Updates,” Mar. 2021.

[211] J. Orlovska, C. Wickman, and R. Söderberg, “Capturing Customer Profile Enables in-Vehicle User Identification: Design for Data-Based User Behavior Evaluation,” in *Smart Innovation, Systems and Technologies.* Springer Singapore, 2019, pp. 665–675.

[212] J. Orlovska, F. Novakazi, C. Wickman, and R. Soderberg, “Mixed-Method Design for User Behavior Evaluation of Automated Driver Assistance Systems: An Automotive Industry Case,” *Proceedings of the Design Society: International Conference on Engineering Design*, vol. 1, no. 1, pp. 1803–1812, Jul. 2019.

[213] Q. Yang, A. Scuito, J. Zimmerman, J. Forlizzi, and A. Steinfeld, “Investigating How Experienced UX Designers Effectively Work with Machine Learning,” in *Proceedings of the 2018 on Designing Interactive Systems Conference 2018 - DIS ’18.* ACM Press, 2018.

[214] A. Cooper, *The Inmates Are Running the Asylum.* Indianapolis, IN: Sams, 1999.

[215] J. Salminen, S.-g. Jung, and B. Jansen, “The Future of Data-driven Personas: A Marriage of Online Analytics Numbers and Human Attributes,” in *Proceedings of the 21st International Conference on Enterprise Information Systems.* SCITEPRESS - Science and Technology Publications, 2019.

[216] J. Brickey, S. Walczak, and T. Burgess, “Comparing Semi-Automated Clustering Methods for Persona Development,” *IEEE Transactions on Software Engineering*, vol. 38, no. 3, pp. 537–546, May 2012.

[217] J. J. McGinn and N. Kotamraju, “Data-driven persona development,” in *Proceeding of the Twenty-Sixth Annual CHI Conference on Human Factors in Computing Systems - CHI ’08.* ACM Press, 2008.

[218] X. Zhang, H.-F. Brown, and A. Shankar, “Data-driven Personas,” in *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems.* ACM, May 2016.

[219] S.-G. Jung, J. An, H. Kwak, M. Ahmad, L. Nielsen, and B. J. Jansen, “Persona Generation from Aggregated Social Media Data,” in *Proceedings of the 2017 CHI Conference Extended Abstracts on Human Factors in Computing Systems.* ACM, May 2017.

[220] J. An, H. Kwak, S. Jung, J. Salminen, M. Admad, and B. Jansen, “Imaginary People Representing Real Numbers,” *ACM Transactions on the Web*, vol. 12, no. 4, pp. 1–26, Nov. 2018.

[221] J. Orlovska, C. Wickman, and R. Söderberg, "Naturalistic driving study for Automated Driver Assistance Systems (ADAS) evaluation in the Chinese, Swedish and American markets." *Procedia CIRP*, vol. 93, pp. 1286–1291, 2020.

[222] P. Chaovalit, C. Saiprasert, and T. Pholprasit, "A method for driving event detection using SAX on smartphone sensors," in *2013 13th International Conference on ITS Telecommunications (ITST)*. IEEE, Nov. 2013.

[223] S. Daptardar, V. Lakshminarayanan, S. Reddy, S. Nair, S. Sahoo, and P. Sinha, "Hidden Markov Model based driving event detection and driver profiling from mobile inertial sensor data," in *2015 IEEE SENSORS*. IEEE, Nov. 2015.

[224] B. Bose, J. Dutta, S. Ghosh, P. Pramanick, and S. Roy, "D&RSense: Detection of Driving Patterns and Road Anomalies," in *2018 3rd International Conference On Internet of Things: Smart Innovation and Usages (IoT-SIU)*. IEEE, Feb. 2018.

[225] D. Mitrovic, "Reliable Method for Driving Events Recognition," *IEEE Transactions on Intelligent Transportation Systems*, vol. 6, no. 2, pp. 198–205, Jun. 2005.

[226] P. Leakkaw and S. Panichpapiboon, "Real-Time Lane Change Detection Through Steering Wheel Rotation," in *2018 IEEE Vehicular Networking Conference (VNC)*. IEEE, Dec. 2018.

[227] M. V. Ly, S. Martin, and M. M. Trivedi, "Driver classification and driving style recognition using inertial sensors," in *2013 IEEE Intelligent Vehicles Symposium (IV)*. IEEE, Jun. 2013.

[228] G. Bailly, A. Oulasvirta, T. Kötzing, and S. Hoppe, "MenuOptimizer," in *Proceedings of the 26th Annual ACM Symposium on User Interface Software and Technology - UIST '13*. ACM Press, 2013.

[229] P. Green, T. Kang, and B. Lin, "Touch Screen Task Element Times for Improving SAE Recommended Practice J2365: First Proposal," 2015.

[230] H. Kim and H. Song, "Evaluation of the safety and usability of touch gestures in operating in-vehicle information systems with visual occlusion," *Applied Ergonomics*, vol. 45, no. 3, pp. 789–798, May 2014.

[231] M. Pettitt and G. Burnett, "Visual Demand Evaluation Methods for In-Vehicle Interfaces," *International Journal of Mobile Human Computer Interaction*, vol. 2, no. 4, pp. 45–57, Oct. 2010.

[232] D. Tang, A. Agarwal, D. O'Brien, and M. Meyer, "Overlapping experiment infrastructure," in *Proceedings of the 16th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining - KDD '10*. ACM Press, 2010.

[233] R. Ros and P. Runeson, "Continuous experimentation and A/B testing," in *Proceedings of the 4th International Workshop on Rapid Continuous Software Engineering*. ACM, May 2018.

[234] F. Giaimo, H. Andrade, and C. Berger, "The Automotive Take on Continuous Experimentation: A Multiple Case Study," in *2019 45th Euromicro Conference on Software Engineering and Advanced Applications (SEAA)*. Los Alamitos, CA, USA: IEEE Computer Society, Aug. 2019, pp. 126–130.

[235] A. Marrella and T. Catarci, "Measuring the Learnability of Interactive Systems Using a Petri Net Based Approach," in *Proceedings of the 2018 Designing Interactive Systems Conference*. ACM, Jun. 2018.

[236] I. Gerostathopoulos, S. Kugele, C. Segler, T. Bures, and A. Knoll, "Automated Trainability Evaluation for Smart Software Functions," in *2019 34th IEEE/ACM International Conference on Automated Software Engineering (ASE)*. IEEE, Nov. 2019.

[237] S. Fox, K. Karnawat, M. Mydland, S. Dumais, and T. White, "Evaluating implicit measures to improve web search," *ACM Transactions on Information Systems*, vol. 23, no. 2, pp. 147–168, Apr. 2005.

[238] F. Lachner, F. Fincke, and A. Butz, "UX Metrics: Deriving Country-Specific Usage Patterns of a Website Plug-In from Web Analytics," in *Human-Computer Interaction – INTERACT 2017*. Springer International Publishing, 2017, pp. 142–159.

[239] P. Hebron, *Machine Learning for Designers*. Sebastopol, CA: O'Reilly Media, 2016.

[240] S. Carter and M. Nielsen, "Using Artificial Intelligence to Augment Human Intelligence," *Distill*, vol. 2, no. 12, Dec. 2017.

[241] P. Hohl, J. Münch, K. Schneider, and M. Stupperich, "Forces that Prevent Agile Adoption in the Automotive Domain," in *Product-Focused Software Process Improvement*. Springer International Publishing, 2016, pp. 468–476.

[242] B. Katumba and E. Knauss, "Agile Development in Automotive Software Development: Challenges and Opportunities," in *Product-Focused Software Process Improvement*. Springer International Publishing, 2014, pp. 33–47.

[243] Safety and Human Factors Standards Steering Committee, "Navigation and Route Guidance Function Accessibility While Driving," SAE International, Tech. Rep.

[244] B. Still and K. Crane, *Fundamentals of User-Centered Design: A Practical Approach*, 1st ed. Boca Raton: CRC Press, Feb. 2017.

[245] C. Abras, D. Maloney-Krichmar, and J. Preece, "User-Centered Design," in *Encyclopedia of Human-Computer Interaction*. Thousand Oaks, CA, USA: Sage Publications, 2004.

[246] K.-H. Renner and N.-C. Jacob, *Das Interview: Grundlagen und Anwendung in Psychologie und Sozialwissenschaften*, ser. Basiswissen Psychologie. Berlin, Heidelberg: Springer Berlin Heidelberg, 2020.

[247] M. Broy, I. H. Kruger, A. Pretschner, and C. Salzmann, "Engineering Automotive Software," *Proceedings of the IEEE*, vol. 95, no. 2, pp. 356–373, Feb. 2007.

[248] S. McLellan, A. Muddimer, and S. C. Peres, "The Effect of Experience on System Usability Scale Ratings," *Journal of Usability Studies*, vol. 7, no. 2, pp. 56–67, Feb. 2012.

[249] D. Albers, J. Radlmayr, A. Loew, S. Hergeth, F. Naujoks, A. Keinath, and K. Bengler, "Usability Evaluation - Advances in Experimental Design in the Context of Automated Driving Human-Machine Interfaces," *Information*, vol. 11, no. 5, p. 240, Apr. 2020.

[250] S. Borsci, S. Federici, S. Bacci, M. Gnaldi, and F. Bartolucci, "Assessing User Satisfaction in the Era of User Experience: Comparison of the SUS, UMUX, and UMUX-LITE as a Function of Product Experience," *International Journal of Human-Computer Interaction*, vol. 31, no. 8, pp. 484–495, Aug. 2015.

[251] J. R. Lewis and J. Sauro, "Item Benchmarks for the System Usability Scale," *Journal of Usability Studies*, vol. 13, no. 3, pp. 158–167, May 2018.

[252] B. Cleland, J. Wallace, R. Bond, S. Muuraiskangas, J. Pajula, G. Epelde, M. Arrúe, R. Álvarez, M. Black, M. D. Mulvenna, D. Rankin, and P. Carlin, "Usability Evaluation of a Co-created Big Data Analytics Platform for Health Policy-Making," in *Human Interface and the Management of Information. Visual Information and Knowledge Management*, S. Yamamoto and H. Mori, Eds. Cham: Springer International Publishing, 2019, vol. 11569, pp. 194–207.

[253] D. A. Magezi, "Linear mixed-effects models for within-participant psychology experiments: An introductory tutorial and free, graphical user interface (LMMgui)," *Frontiers in Psychology*, vol. 6, p. 2, Jan. 2015.

[254] R Core Team, *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria, 2022.

[255] D. Bates, M. Mächler, B. Bolker, and S. Walker, "Fitting Linear Mixed-Effects Models Using **lme4**," *Journal of Statistical Software*, vol. 67, no. 1, pp. 1–48, 2015.

[256] A. Kuznetsova, P. B. Brockhoff, and R. H. B. Christensen, "lmerTest Package: Tests in Linear Mixed Effects Models," *Journal of Statistical Software*, vol. 82, no. 13, pp. 1–26, 2017.

[257] R. V. Lenth, "Emmeans: Estimated marginal means, aka least-squares means," 2022.

[258] M. Hlavac, *Stargazer: Well-formatted Regression and Summary Statistics Tables*, Social Policy Institute, Bratislava, Slovakia, 2022.

[259] S. G. Luke, "Evaluating significance in linear mixed-effects models in R," *Behavior Research Methods*, vol. 49, no. 4, pp. 1494–1502, Aug. 2017.

[260] J. W. Tukey, "Comparing Individual Means in the Analysis of Variance," *Biometrics*, vol. 5, no. 2, p. 99, Jun. 1949.

[261] A. Morando, P. Gershon, B. Mehler, and B. Reimer, "A model for naturalistic glance behavior around Tesla Autopilot disengagements," *Accident Analysis & Prevention*, vol. 161, p. 106348, Oct. 2021.

[262] A.-S. Wikman, T. Nieminen, and H. Summala, "Driving experience and time-sharing during in-car tasks on roads of different width," *Ergonomics*, vol. 41, no. 3, pp. 358–372, Mar. 1998.

[263] F. Naujoks, C. Purucker, and A. Neukum, "Secondary task engagement and vehicle automation – Comparing the effects of different automation levels in an on-road experiment," *Transportation Research Part F: Traffic Psychology and Behaviour*, vol. 38, pp. 67–82, Apr. 2016.

[264] S. Nakagawa and H. Schielzeth, "A general and simple method for obtaining R^2 from generalized linear mixed-effects models," *Methods in Ecology and Evolution*, vol. 4, no. 2, pp. 133–142, Feb. 2013.

[265] J. D. Lee, "Dynamics of Driver Distraction: The process of engaging and disengaging," *Annals of Advances in Automotive Medicine. Association for the Advancement of Automotive Medicine. Annual Scientific Conference*, vol. 58, pp. 24–32, 2014.

[266] S. M. Lundberg and S.-I. Lee, "A unified approach to interpreting model predictions," in *Proceedings of the 31st International Conference on Neural Information Processing Systems*, ser. NIPS'17. Red Hook, NY, USA: Curran Associates Inc., 2017, pp. 4768–4777.

[267] A. Barredo Arrieta, N. Díaz-Rodríguez, J. Del Ser, A. Bennetot, S. Tabik, A. Barbado, S. Garcia, S. Gil-Lopez, D. Molina, R. Benjamins, R. Chatila, and F. Herrera, "Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI," *Information Fusion*, vol. 58, pp. 82–115, Jun. 2020.

[268] Q. V. Liao, D. Gruen, and S. Miller, "Questioning the AI: Informing Design Practices for Explainable AI User Experiences," in *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. Honolulu HI USA: ACM, Apr. 2020, pp. 1–15.

[269] G. Wiegand, M. Eiband, M. Haubelt, and H. Hussmann, ""I'd like an Explanation for That!"Exploring Reactions to Unexpected Autonomous Driving," in *22nd International Conference on Human-Computer Interaction with Mobile Devices and Services*. Oldenburg Germany: ACM, Oct. 2020, pp. 1–11.

[270] C. Wang and P. An, "A Mobile Tool that Helps Nonexperts Make Sense of Pretrained CNN by Interacting with Their Daily Surroundings," in *Adjunct Publication of the 23rd International Conference on Mobile Human-Computer Interaction*. Toulouse & Virtual France: ACM, Sep. 2021, pp. 1–5.

[271] F. Doshi-Velez and B. Kim, "Towards a rigorous science of interpretable machine learning," *arXiv preprint arXiv:1702.08608*, 2017.

[272] D. Shin, "The effects of explainability and causability on perception, trust, and acceptance: Implications for explainable AI," *International Journal of Human-Computer Studies*, vol. 146, p. 102551, Feb. 2021.

[273] Z. C. Lipton, "The mythos of model interpretability," *Queue*, vol. 16, no. 3, pp. 31–57, Jun. 2018.

[274] T. Verma, C. Lingenfelder, and D. Klakow, "Defining explanation in an AI context," in *Proceedings of the Third BlackboxNLP Workshop on Analyzing and Interpreting Neural Networks for NLP*. Association for Computational Linguistics, 2020.

[275] Y. Zhang, Q. V. Liao, and R. K. E. Bellamy, "Effect of confidence and explanation on accuracy and trust calibration in AI-assisted decision making," in *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency*. Barcelona Spain: ACM, Jan. 2020, pp. 295–305.

[276] L. S. Shapley, "A value for n-Person games," in *Contributions to the Theory of Games (AM-28), Volume II*. Princeton University Press, Dec. 1953, pp. 307–318.

[277] C. Molnar, *Interpretable Machine Learning*. Lulu.com, Feb. 2020.

[278] S. M. Lundberg, G. Erion, H. Chen, A. DeGrave, J. M. Prutkin, B. Nair, R. Katz, J. Himmelfarb, N. Bansal, and S.-I. Lee, "From local explanations to global understanding with explainable AI for trees," *Nature Machine Intelligence*, vol. 2, no. 1, pp. 56–67, Jan. 2020.

[279] M. T. Ribeiro, S. Singh, and C. Guestrin, ""Why Should I Trust You?"," in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. ACM, Aug. 2016.

[280] S. M. Lundberg, G. G. Erion, and S.-I. Lee, "Consistent individualized feature attribution for tree ensembles," *arXiv preprint arXiv:1802.03888*, 2018.

[281] K. Custer, "100-Car Data," 2018.

[282] J. Bärgman, V. Lisovskaja, T. Victor, C. Flannagan, and M. Dozza, "How does glance behavior influence crash and injury risk? A 'what-if' counterfactual simulation using crashes and near-crashes from SHRP2," *Transportation Research Part F: Traffic Psychology and Behaviour*, vol. 35, pp. 152–169, Nov. 2015.

[283] R. Kohavi, "A study of cross-validation and bootstrap for accuracy estimation and model selection," in *Proceedings of the 14th International Joint Conference on Artificial Intelligence - Volume 2*, ser. IJCAI'95. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc., 1995, pp. 1137–1143.

[284] O. Carsten, F. C. H. Lai, Y. Barnard, A. H. Jamson, and N. Merat, "Control task substitution in semiautomated driving," *Human Factors: The Journal of the Human Factors and Ergonomics Society*, vol. 54, no. 5, pp. 747–761, Sep. 2012.

[285] C. P. Janssen, S. J. Gould, S. Y. Li, D. P. Brumby, and A. L. Cox, "Integrating knowledge of multitasking and interruptions across different perspectives and research methods," *International Journal of Human-Computer Studies*, vol. 79, pp. 1–5, Jul. 2015.

[286] J. O. Wobbrock and J. A. Kientz, "Research contributions in human-computer interaction," *Interactions*, vol. 23, no. 3, pp. 38–44, Apr. 2016.

[287] K.-J. Stol and B. Fitzgerald, "The ABC of Software Engineering Research," *ACM Transactions on Software Engineering and Methodology*, vol. 27, no. 3, pp. 1–51, Sep. 2018.

[288] H. Kim and J. L. Gabbard, "Assessing Distraction Potential of Augmented Reality Head-Up Displays for Vehicle Drivers," *Human Factors: The Journal of the Human Factors and Ergonomics Society*, vol. 64, no. 5, pp. 852–865, Aug. 2022.

[289] M. Berger, P. Ebel, D. Dey, A. Dandekar, B. Barati, B. Pfleging, and R. Bernhaupt, "Together Distracted? The Effect of Driver-Passenger Collaboration on Workload, Glance Behavior, and Driving Performance," in *14th International Conference on Automotive User Interfaces and Interactive Vehicular Applications.* Seoul Republic of Korea: ACM, Sep. 2022, pp. 67–72.

[290] F. Zhang and S. C. Roberts, "Factors affecting drivers' off-road glance behavior while interacting with in-vehicle voice interfaces," *Accident Analysis & Prevention*, vol. 179, p. 106883, Jan. 2023.

[291] T. Kujala and H. Grahn, "Visual Distraction Effects of In-Car Text Entry Methods: Comparing Keyboard, Handwriting and Voice Recognition," in *Proceedings of the 9th International Conference on Automotive User Interfaces and Interactive Vehicular Applications.* Oldenburg Germany: ACM, Sep. 2017, pp. 1–10.

[292] L. Graichen, M. Graichen, and J. F. Krems, "Evaluation of Gesture-Based In-Vehicle Interaction: User Experience and the Potential to Reduce Driver Distraction," *Human Factors: The Journal of the Human Factors and Ergonomics Society*, vol. 61, no. 5, pp. 774–792, Aug. 2019.

[293] D. Stiegemeier, S. Bringeland, J. Kraus, and M. Baumann, "User Experience of In-Vehicle Gesture Interaction: Exploring the Effect of Autonomy and Competence in a Mock-Up Experiment," in *Proceedings of the 14th International Conference on Automotive User Interfaces and Interactive Vehicular Applications.* Seoul Republic of Korea: ACM, Sep. 2022, pp. 285–296.

[294] T. Hastie, R. Tibshirani, and J. H. Friedman, *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, 2nd ed., ser. Springer Series in Statistics. New York, NY: Springer, 2009.

[295] D. V. Carvalho, E. M. Pereira, and J. S. Cardoso, "Machine Learning Interpretability: A Survey on Methods and Metrics," *Electronics*, vol. 8, no. 8, p. 832, Jul. 2019.

[296] H. Gwyther and C. Holland, "The effect of age, gender and attitudes on self-regulation in driving," *Accident Analysis & Prevention*, vol. 45, pp. 19–28, Mar. 2012.

[297] L. K. Donorfio, L. A. D'Ambrosio, J. F. Coughlin, and M. Mohyde, "Health, safety, self-regulation and the older driver: It's not just a matter of age," *Journal of Safety Research*, vol. 39, no. 6, pp. 555–561, Jan. 2008.

[298] H.-Y. W. Chen and B. Donmez, "What drives technology-based distractions? A structural equation model on social-psychological factors of technology-based driver distraction engagement," *Accident Analysis & Prevention*, vol. 91, pp. 166–174, Jun. 2016.

[299] M. Regan, "Driver distraction: Reflections on the past, present and future," *Distracted driving. Sydney, NSW: Australasian College of Road Safety*, pp. 29–73, 2007.

[300] M. Kim, E. Seong, Y. Jwa, J. Lee, and S. Kim, "A Cascaded Multimodal Natural User Interface to Reduce Driver Distraction," *IEEE Access*, vol. 8, pp. 112 969–112 984, 2020.

[301] B. Pfleging, S. Schneegass, and A. Schmidt, "Multimodal interaction in the car," in *Proceedings of the 4th International Conference on Automotive User Interfaces and Interactive Vehicular Applications - AutomotiveUI '12*. ACM Press, 2012.

[302] V. Ramanishka, Y.-T. Chen, T. Misu, and K. Saenko, "Toward Driving Scene Understanding: A Dataset for Learning Driver Behavior and Causal Reasoning," 2018.

[303] Y. Bengio, A. Courville, and P. Vincent, "Representation Learning: A Review and New Perspectives," Apr. 2014.

[304] S. J. Gershman, E. J. Horvitz, and J. B. Tenenbaum, "Computational rationality: A converging paradigm for intelligence in brains, minds, and machines," *Science*, vol. 349, no. 6245, pp. 273–278, Jul. 2015.

[305] R. L. Lewis, A. Howes, and S. Singh, "Computational Rationality: Linking Mechanism and Behavior Through Bounded Utility Maximization," *Topics in Cognitive Science*, vol. 6, no. 2, pp. 279–311, Apr. 2014.

[306] A. Oulasvirta, J. P. P. Jokinen, and A. Howes, "Computational Rationality as a Theory of Interaction," in *CHI Conference on Human Factors in Computing Systems*. New Orleans LA USA: ACM, Apr. 2022, pp. 1–14.

[307] J. P. P. Jokinen and T. Kujala, "Modelling Drivers' Adaptation to Assistance Systems," in *13th International Conference on Automotive User Interfaces and Interactive Vehicular Applications*. Leeds United Kingdom: ACM, Sep. 2021, pp. 12–19.

[308] J. P. P. Jokinen, T. Kujala, and A. Oulasvirta, "Multitasking in Driving as Optimal Adaptation Under Uncertainty," *Human Factors: The Journal of the Human Factors and Ergonomics Society*, vol. 63, no. 8, pp. 1324–1341, Dec. 2021.

Erklärung zur Dissertation

"Hiermit versichere ich an Eides statt, dass ich die vorliegende Dissertation selbstständig und ohne die Benutzung anderer als der angegebenen Hilfsmittel und Literatur angefertigt habe. Alle Stellen, die wörtlich oder sinngemäß aus veröffentlichten und nicht veröffentlichten Werken dem Wortlaut oder dem Sinn nach entnommen wurden, sind als solche kenntlich gemacht. Ich versichere an Eides statt, dass diese Dissertation noch keiner anderen Fakultät oder Universität zur Prüfung vorgelegen hat; dass sie - abgesehen von unten angegebenen Teilpublikationen und eingebundenen Artikeln und Manuskripten - noch nicht veröffentlicht worden ist sowie, dass ich eine Veröffentlichung der Dissertation vor Abschluss der Promotion nicht ohne Genehmigung des Promotionsausschusses vornehmen werde. Die Bestimmungen dieser Ordnung sind mir bekannt. Darüber hinaus erkläre ich hiermit, dass ich die Ordnung zur Sicherung guter wissenschaftlicher Praxis und zum Umgang mit wissenschaftlichem Fehlverhalten der Universität zu Köln gelesen und sie bei der Durchführung der Dissertation zugrundeliegenden Arbeiten und der schriftlich verfassten Dissertation beachtet habe und verpflichte mich hiermit, die dort genannten Vorgaben bei allen wissenschaftlichen Tätigkeiten zu beachten und umzusetzen. Ich versichere, dass die eingereichte elektronische Fassung der eingereichten Druckfassung vollständig entspricht."

Teilpublikationen:

[1] P. Ebel, F. Brokhausen, and A. Vogelsang, "The Role and Potentials of Field User Interaction Data in the Automotive UX Development Lifecycle: An Industry Perspective," in *Proceedings of the 12th International Conference on Automotive User Interfaces and Interactive Vehicular Applications.* Virtual Event DC USA: ACM, Sep. 2020, pp. 141–150

[2] P. Ebel, J. Orlovska, S. Hünemeyer, C. Wickman, A. Vogelsang, and R. Söderberg, "Automotive UX design and data-driven development: Narrowing the gap to support practitioners," *Transportation Research Interdisciplinary Perspectives*, vol. 11, p. 100455, Sep. 2021

[3] P. Ebel, C. Lingenfelder, and A. Vogelsang, "Visualizing Event Sequence Data for User Behavior Evaluation of In-Vehicle Information Systems," in *Proceedings of the 13th International Conference on Automotive User Interfaces and Interactive Vehicular Applications.* Leeds United Kingdom: ACM, Sep. 2021, pp. 219–229

[4] P. Ebel, M. Berger, C. Lingenfelder, and A. Vogelsang, "How Do Drivers Self-Regulate their Secondary Task Engagements? The Effect of Driving Automation on Touchscreen Interactions and Glance Behavior," in *Proceedings of the 14th International Conference on Automotive User Interfaces and Interactive Vehicular Applications.* Seoul Republic of Korea: ACM, Sep. 2022, pp. 263–273

[5] P. Ebel, K. J. Gülle, C. Lingenfelder, and A. Vogelsang, "ICEBOAT: An Interactive User Behavior Analysis Tool for Automotive User Interfaces," in *The Adjunct Publication of the 35th Annual ACM Symposium on User Interface Software and Technology*, Aug. 2022

[6] P. Ebel, C. Lingenfelder, and A. Vogelsang, “On the forces of driver distraction: Explainable predictions for the visual demand of in-vehicle touchscreen interactions,” *Accident Analysis & Prevention*, vol. 183, p. 106956, Apr. 2023

[7] P. Ebel, C. Lingenfelder, and A. Vogelsang, “Multitasking while driving: How drivers self-regulate their interaction with in-vehicle touchscreens in automated driving,” *International Journal of Human–Computer Interaction*, pp. 1–18, 2023

[8] P. Ebel, K. J. Gülle, C. Lingenfelder, and A. Vogelsang, “Exploring Millions of User Interactions with ICEBOAT: Big Data Analytics for Automotive User Interfaces,” in *Proceedings of the 15th International Conference on Automotive User Interfaces and Interactive Vehicular Applications*, Ingolstadt, Germany, 2023

10.08.2023	Patrick Ebel
Datum	**Name**

Patrick Ebel

Unterschrift

Zugänglichkeit von Daten und Materialien

Die Daten und Materialien die zur Reproduktion der in der Dissertation dargestellten Ergebnisse benötigt werden, wurden zum Teil von der MBition GmbH zur Verfügung gestellt und sind gemäß des Zusammenarbeitsvertrags der Universität zu Köln und der MBition GmbH vertraulich (vgl. §11. des Zusammenarbeitsvertrags). Diese Daten sind bei der MBition GmbH gelagert. Im Falle eines Verfahrens im Sinne der guten wissenschaftlichen Praxis und/oder des Verdachts einer Täuschung (vgl. §18 Abs. 1 der Promotionsordnung der Mathematisch-Naturwissenschaftlichen Fakultät der Universität zu Köln vom 12. März 2020) regelt §11 Abs. 2 des Zusammenarbeitsvertrags die Aufhebung der Vertraulichkeitsverpflichtung. Weitere Daten, die nicht im Zusammenhang mit der MBition GmbH stehen, werden auf einem Datenserver der Arbeitsgruppe "Software and Systems Engineering" von Prof. Dr. Vogelsang gespeichert.