RESEARCH-ARTICLE

Computational Models for In-Vehicle User Interface Design: A Systematic Literature Review

MARTIN LORENZ, University of Leipzig, Leipzig, Sachsen, Germany

TIAGO AMORIM, University of Cologne, Koln, Nordrhein-Westfalen, Germany

DEBARGHA (DAVE) DEY, Cornell Tech, New York, NY, United States

MERSEDEH SADEGHI, University of Cologne, Koln, Nordrhein-Westfalen, Germany

PATRICK EBEL, University of Leipzig, Leipzig, Sachsen, Germany

Published: 22 September 2024

Citation in BibTeX format

AutomotiveUI '24: 16th International Conference on Automotive User Interfaces and Interactive Vehicular Applications
September 22 - 25, 2024
CA, Stanford, USA

Conference Sponsors:
SIGCHI

AutomotiveUI '24: Proceedings of the 16th International Conference on Automotive User Interfaces and Interactive Vehicular Applications (September 2024)
https://doi.org/10.1145/3640792.3675735
ISBN: 9798400705106

Computational Models for In-Vehicle User Interface Design: A Systematic Literature Review

Martin Lorenz
lorenz@cs.uni-leipzig.de
ScaDS.AI, Leipzig University
Leipzig, Germany

Tiago Amorim
amorim@cs.uni-koeln.de
University of Cologne
Cologne, Germany

Debargha Dey
debargha.dey@cornell.edu
Cornell Tech
New York, USA

Mersedeh Sadeghi
sadeghi@cs.uni-koeln.de
University of Cologne
Cologne, Germany

Patrick Ebel
ebel@uni-leipzig.de
ScaDS.AI, Leipzig University
Leipzig, Germany

ABSTRACT

In this review, we analyze the current state of the art of computational models for in-vehicle User Interface (UI) design. Driver distraction, often caused by drivers performing Non Driving Related Tasks (NDRTs), is a major contributor to vehicle crashes. Accordingly, in-vehicle UIs must be evaluated for their distraction potential. Computational models are a promising solution to automate this evaluation, but are not yet widely used, limiting their real-world impact. We systematically review the existing literature on computational models for NDRTs to analyze why current approaches have not yet found their way into practice. We found that while many models are intended for UI evaluation, they focus on small and isolated phenomena that are disconnected from the needs of automotive UI designers. In addition, very few approaches make predictions detailed enough to inform current design processes. Our analysis of the state of the art, the identified research gaps, and the formulated research potentials can guide researchers and practitioners toward computational models that improve the automotive UI design process.

CCS CONCEPTS

• **Human-centered computing → HCI theory, concepts and models**; **User models**; **User studies**; • **General and reference → Surveys and overviews**.

KEYWORDS

computational modeling, user interfaces, literature review, cognitive modeling, automated driving

ACM Reference Format:
Martin Lorenz, Tiago Amorim, Debargha Dey, Mersedeh Sadeghi, and Patrick Ebel. 2024. Computational Models for In-Vehicle User Interface Design: A Systematic Literature Review. In *16th International Conference on Automotive User Interfaces and Interactive Vehicular Applications (AutomotiveUI '24), September 22–25, 2024, Stanford, CA, USA.* ACM, New York, NY, USA, 12 pages. https://doi.org/10.1145/3640792.3675735

AutomotiveUI '24, September 22–25, 2024, Stanford, CA, USA

ACM ISBN 979-8-4007-0510-6/24/09
https://doi.org/10.1145/3640792.3675735

1 INTRODUCTION

Driver distraction is a leading cause of vehicle crashes [12, 28, 96]. Nevertheless, modern vehicles are equipped with a variety of in-vehicle UIs that allow the driver to access and control driving and infotainment functions while driving. While interacting with such in-vehicle UIs, drivers divide their attention between the primary driving task and a secondary Non Driving Related Tasks (NDRTs). This diversion of attention away from the road can impair driving performance and increase crash risk. While driver distraction, its consequences, and mitigation strategies have been discussed for many years [96], the discussion has recently gained momentum due to large center stack touchscreens becoming the de facto standard interface [21, 22, 63, 87, 89]. Growing concern about the impact of touchscreens and infotainment displays on driver distraction has prompted regulatory bodies like the European New Car Assessment Programme (Euro NCAP) to mandate a return to haptic controls for critical functionalities in order for vehicles to achieve a five-star safety rating [4]. Currently, the automotive industry primarily relies on empirical user studies to evaluate in-vehicle interfaces for usability and distraction potential. However, the complexity of modern infotainment systems makes these studies resource-intensive and limits their ability to provide holistic distraction evaluations [16]. A promising solution are computational models that can automate (parts of) the interface evaluation and thus help designers to evaluate designs not only more efficiently, but also earlier in the design process and on a larger scale. Outside the automotive domain, there are already many approaches that make detailed predictions about the impact of design features on user behavior, for example in web design [5], mobile app design [11], or desktop GUI design [66]. However, while computational models for in-vehicle interface evaluation exist [73, 74], they have not yet found their way into the industrial UX design process [19]. Understanding what models currently exist, what their capabilities are, and what is missing could help in the challenge of making computational models a standard tool within the design process of in-vehicle UIs.

In this systematic literature review, therefore, we analyze the current state of computational models that can be used to support the automotive UIs design process. In particular, we focus on the relationship between the task being modeled, the model being used, and the level of detail in the modeling approach.

Contribution Statement: The contribution of this paper is two-fold: we (1) present an overview of the state of the art of computational models for evaluating NDRTs with in-vehicle UIs and (2) identify current research gaps and future research potentials. In doing so, we provide the first systematic review of computational models for in-vehicle UI design.

The identified research gaps and potentials can help researchers and practitioners alike to develop models and tools that improve the evaluation and design of in-vehicle UIs in the automotive industry, with a direct impact on the systems that are on our roads.

2 BACKGROUND AND RELATED WORK

In this section, we introduce the idea of using computational models to improve automotive user interfaces. First, we discuss why distraction, workload, and glance behavior are common metrics for evaluating driver interactions with automotive UIs and how these interactions can affect road safety. We then review the role of computational models in Human Computer Interaction (HCI), and draw the connection to the automotive domain. Finally, we explain how our review relates to previous literature.

2.1 Automotive User Interfaces, Driver Distraction, and Road Safety

Automotive user interfaces include all interfaces that allow the driver to interact with the vehicle. The purpose of these interfaces may be to provide *information* to support the driving task or to provide *entertainment* to the driver [26]. Interfaces that provide both, such as large center stack touchscreens, are called *infotainment*-systems [59]. Drivers can interact with such interfaces in a variety of modalities, including but not limited to speech, touch, or hard keys [59]. Recently, touchscreens not only became the de-facto interface to acess most of the functionalities but also a frequently discussed topic when it comes to driver distraction [15, 17, 20, 70]. Driver distraction is defined as "a diversion of attention from the activities critical for safe driving toward a competing activity" [50]. The relevance of touchscreen interactions for driver distraction becomes clear considering that drivers are forced to divide their visual attention between the primary driving task and the secondary touchscreen task. In contrast to hard key interactions, drivers do not receive any haptic feedback and must visually verify that their interaction has resulted in the intended action, which increases the amount of time they spend looking away from the road. If drivers are too engaged in NDRTs their awareness of the driving environment and situation around them decreases, which can be a contributing factor to road accidents[78]. For this reason, driver distraction has to be reduced as, for example, stated in the National Highway Traffic Association (NHTSA) driver distraction guidelines [1]. To do so, vehicle manufactures traditionally conduct extensive user studies to evaluate and reduce the distraction potential of the in-vehicle user interfaces. However, modern systems are too complex and user studies are too time-consuming and expensive, that it is not possible to perform holistic system tests in a sufficient manner. Computational models can provide a remedy here, because they can already be used in the design cycle and not only after developing the UI. They are way less expensive than traditional user studies and they can be used for testing complex scenarios.

2.2 Computational Models

Computational models in the area of HCI can help researchers and practitioners alike to understand, explain, and predict human behavior. In line with Shmueli [84], models can be divided into statistical models, which are concerned with the *generation* of new insights on human behavior and computational models that *predict* or *simulate* human behavior. Although both types of models can theoretically be used for either purpose, in HCI statistical models are used primarily for explanation and computational models for prediction.

Computational models have been used to evaluate driver behavior since the 1970s [71]. Aligned with Park and Zahabi [71], following categories of computational models to model driver behavior exist. For completeness we also introduce Machine Learning (ML) models:

Adaptive Control of Thought–Rational (ACT-R). ACT-R is a cognitive architecture and was developed by Anderson et al. [3]. The model is composed of several modules, representing different cognitive functions, such as vision or memory. ACT-R is used in the automotive domain in multiple ways for driver behavior prediction, such as braking, lane changing, multitasking or steering [71].

Goals, Operations, Methods, and Selection Rules (GOMS). These models describe the relationship between a task and a user knowing how to achieve the task. *Goals* are the goals the user wants to achieve. *Operators* are the basic actions of an application, *methods* describe a sequence of operators and sub-goals the user learned before to accomplish a goal, and *selection rules* are heuristics that differ between users to select a method when multiple methods could be applied to accomplish the goal [39]. The most widely used GOMS variation for modeling drivers secondary task interaction are Keystroke-Level-Models (KLMs) [7]. Such approaches decompose an entire task into a sequence of specific primitive operators and empirically determine the interaction time for each operator. The total task time is then equal to the sum of the individual operator times.

Queueing Networks (QNs). QN-models are models based on the mathematical field of Queueing Theory. A QN consists of several connected nodes (called servers), that provide some service to a so called customer. After processing, the customer follows the network to the next server. A general example for a QN is a routing network for internet traffic. As computational models in the field of HCI those models help to employ some kind of real multitasking behavior, such as in QN-MHP by Liu et al. [57] and Liu [56].

Machine Learning (ML). ML techniques, such as *supervised* or *unsupervised* ML, can learn from data to predict human behavior. Therefore, these algorithms optimize the parameters of a given model. Supervised methods transform input data to labeled output data during training, minimizing a loss-function. Unsupervised methods learn without labeling the output. An assumption for these models is that user behavior can be modeled via distinct clusters [68].

Others. This category contains models such as Saliency, Effort, Expectancy, and task Value model (SEEV) or Improved Performance Research Integration Tool (IMPRINT). SEEV was developed by Wickens et al. [90] and is a cognitive architecture. It connects the elements of attention allocation in dynamic environments, namely the *salience* of an event in an area of interest and the *effort* to look at this event, with the *expected value* such an action may have for the user. [90] IMPRINT is a discrete human-performance modeling software tool developed by the US Army Research Laboratory. Based on four classes of human resources that may contribute to driver workload, namely visual, cognitive, auditory, and psychomotor, it computes demand scores for different tasks. The linking between the task or interfaces and the score has to be done manually [42, 71].

Recent work shows that there is a strong need for computational models that are valuable in practice and make predictions that are explicitly linked to the design of user interfaces [19, 35, 64]. Therefore, in this paper we focus on predictive computational models that can be used to evaluate and inform user interface designs and assist designers in their design process. Outside of the automotive domain, various approaches of such computational models already exist. Li et al. [54] apply Deep Learning methods to predict human performance in various vertical menu selection tasks. Leiva et al. [52] predict how different cohorts perceive web page aesthetics using computational models, and Rawles et al. [75] utilize machine learning methods to generate models of mobile device control for Android based systems. Recent works also employ the concept of *computational rationality* [67]. Therefore, these models use Reinforcement Learning (RL) methods where a control policy is learned automatically overcoming the need for large amounts of labeled training data [61]. Such approaches have been used to model various tasks [61], such as pointing [33], menu search [9] or touchscreen typing [40].

2.3 Related Work

Previous approaches have focused on computational models within the domain of driver behavior modeling. Janssen et al. [36] not only provide a framework that defines the terms *agents, environments*, and *scenarios* along specified dimensions, they also identify research opportunities in the human-vehicle interaction domain. Janssen et al. [36] conclude that there is a need for models and studies, that simulate *either* the environment *or* the scenario, but not both. This would increase the realism compared to simulator studies but would not be as costly or resource intensive as naturalistic driving studies [36].

The work of Park and Zahabi [71] reviews Human Performance Models in the surface transportation domain regarding the prediction of driver behavior and the interaction with NDRTs. Therefore Park and Zahabi [71] analyze a combinations of dependent and independent variables and the used model type, where they identify the following categories: *ACT-R, QNs, GOMS*, and an *others* category with models such as SEEV or IMPRINT. The main purpose of the paper is to help modeling experts to select models appropriate to their modeling needs regarding dependent and independent variables.

In their review from 2011, Hurts et al. [32] investigate the mental demand placed on the driver by primary driving tasks, such as lane keeping, speed control or following behavior, and NDRTs, such as the use of nomadic devices or the control of Heating, Ventilation and Air Conditioning (HVAC). They also review how the distracting effects of various tasks, , such as occlusion tasks, lane changing tasks or car following tasks, can be assessed by measuring driving performance or vision-related measures, such as gaze behavior and what official guidelines exist for such tasks.

The review by Doshi and Trivedi [14] (2011) looks into 19 studies and how these model tactical driving maneuver prediction. The aim of these models is to model short term goals, such as turns, lane changes, and stops. Doshi and Trivedi [14] categorize the respective models by inputs, the environment and scenario of data acquisition, the used modeling algorithm such as CAN data, physiological data, motoric behavior or the driving environment (naturalistic or simulator), in which the data were collected. Furthermore, the authors evaluate the prediction accuracy of the models in scope. They find that “models incorporating measures of driver behavior perform better, and earlier at predicting [driving maneuvers]” [14, p. 1896].

All the existing reviews investigate different kind of models for cognitive load, driver maneuver prediction or which models are the most state of the art modelling a NDRT given an environment. However, none of these models predicts the distraction potential of an In-Vehicle-Information-System (IVIS) while driving.

3 METHOD

The goal of this literature review is to present, map, and compare computational models in the automotive domain that can improve the interaction between the driver and the in-vehicle UI . We report our methods and results according to the PRISMA 2020 checklist [69]. Our identification process consists of a database search and a subsequent round of forward and backward snowballing to further extent the set of papers as suggested by Wohlin [92]. The complete process is visualized in Figure 1.

3.1 Database Search

We searched three different databases, namely Scopus, IEEE Explore, and the ACM Digital Library, to ensure the broadest coverage possible. To extract relevant papers we developed our search string according to the guidelines presented by Kitchenham et al. [43]. Based on the research questions and intended contribution of this work we developed a list of synonyms that represent the fact that we are searching for approaches that predict specific phenomena relevant to the interaction between driver and vehicle. We performed trial searches using different combinations of search strings and evaluated the results according to two criteria, namely the number of papers returned (rationale: *Is it manageable to review this number of papers?*) and the coverage of papers we considered relevant in an initial collection of papers (rationale: *Are we retrieving the papers we are interested in?*). The final search strategy considers only full research articles and applies the keywords to a combination of article title and article abstract. The database search was conducted

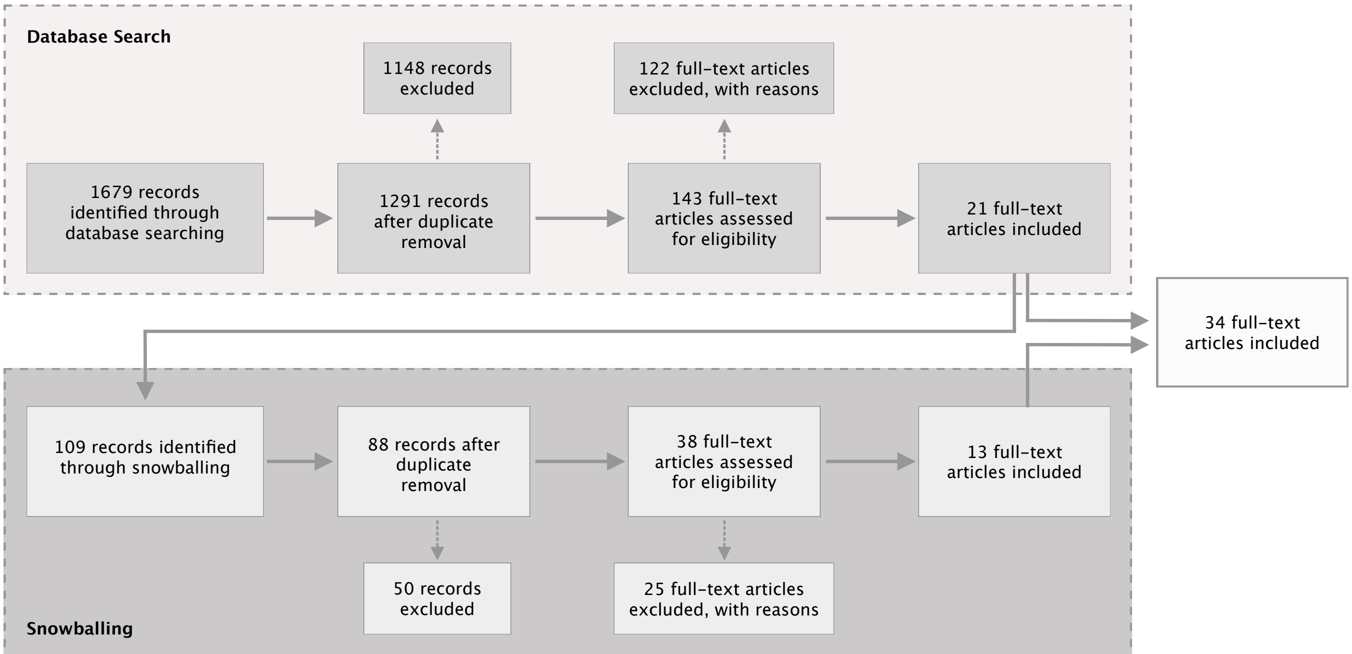

Figure 1: The search process.

in two stages, with the first search on May 4, 2022 and the final search on February 29, 2024. We used the following search string:[1]

> (`Title`: "driver behavior" ∨ "glance behavior" ∨ "driver behaviour" ∨ "glance behaviour" ∨ "distract*" ∨ "secondary task" ∨ "workload" ∨ "visual demand" ∨ "cognitive demand" ∨ "mental demand" ∨ "task demand" ∨ "visual load" ∨ "cognitive load" ∨ "mental load" ∨ "task load" ∨ "visual attention" ∨ "driver attention")
> ∧ (`Title`: "predict*" ∨ "model*" ∨ "assess*" ∨ "detect*" ∨ "simulat*" ∨ "classif*" ∨ "identif*")
> ∧ (`Abstract`: "car" ∨ "automotive" ∨ "vehicle" ∨ "ivis" ∨ "transportation")

3.2 Database Selection Process

The database search described in 3.1 resulted in 1679 papers. For the paper selection process, we used CADIMA [44], a free web-based tool for conducting collaborative systematic reviews. After removing all duplicates using the semi-automated duplicate removal tool provided by CADIMA, we selected relevant papers in a two-stage process. In the first stage, after an initial consistency check in which 100 papers were independently reviewed by three authors, each paper was coded for eligibility based on whether its title and abstract met inclusion criteria IC1 and IC2 (described below). To ensure consistency, each paper was coded by two independent reviewers, and differences were discussed in conflict resolution meetings. At the title and abstract stage, 1148 articles were excluded. The full-texts of the remaining 143 articles were then coded for all inclusion and exclusion criteria as described in detail the codebook and listed below:

Inclusion Criteria:

- IC1: The approach develops/evaluates/validates/utilizes a computational model of any kind.
- IC2: The approach is concerned with the interaction between the driver and the IVIS or another (digital) mobile device.
- IC3: The presented model quantifies the effect of secondary task engagements on driver behavior or vice versa. This excludes approaches that only predict a certain driver state (e.g., distracted driving).

Exclusion Criteria:

- EC1: The paper was not subject of a peer review process.
- EC2: The model is purely theoretical/conceptual.
- EC3: No evaluation is presented.
- EC4: No details regarding the modeling (method/algorithm, inputs, outputs) are presented.
- EC5: No details regarding the data used to either evaluate or fit/train the model is given.

Again, each article was coded by two coders and differences were discussed in conflict resolution meetings. We also excluded 5 papers because the full-text was either not accessible or not assessable (mostly because the paper was written in a language other than English) or because a duplicate was not detected in previous stages. After the full-text stage, 26 articles were included for further analysis. The database selection process resulted in 21 papers.

3.3 Forward and Backward Snowballing

Based on the final set of 21 papers from the database search, we performed a round of forward and backward snowballing to further extend our results. We performed the snowballing according to the guidelines suggested by [92]. In the backward snowballing process,

[1]The asterisk (*) indicates a wildcard character and the he search queries specific to each of the databases are given in the supplemental material: https://osf.io/6dytr

we examined the reference lists of the 21 papers resulting from the database search. In the forward snowballing process, we examined all papers that cited the 21 papers according to Google Scholar. If the title of a paper appeared of interested we included the paper for further examination. This process was performed by two coders and the resulting papers were merged. After removing duplicates, 88 candidate papers were included for further review. Forward and backward snowballing was performed on April 13, 2024.

3.4 Snowballing Selection Process

Methodologically, the snowballing selection process was the same as the database selection process described in subsection 3.2. After reviewing the titles and abstracts, the list of papers was reduced from 88 to 38, resulting in 13 papers after full-text review. These 13 papers were then added to the 21 papers, resulting in a final list of 34 papers for data extraction. During both selection processes, the co-authors performed between 349 and 1031 title and abstract reviews and between 67 and 135 full-text reviews.

3.5 Data Extraction Process

The data extraction process aims to identify features within the papers that are relevant to answering our research questions. To achieve this, we use a mixture of a priori and inductive coding [76].

We categorized the models across 8 dimensions and several codes per dimension. The set of 8 distinct dimensions (TARGET VARIABLE, MODALITY, SYSTEM INTERACTION, ENVIRONMENT, SCENARIO, MODEL INPUTS, INTENDED APPLICATION, MODEL TYPE) was developed a priori based on the expertise of the co-authors and an initial screening of papers published in a related Dagstuhl Report [36]. The dimensions are reported in Table 1, and served as the starting point for the codes, which we generated using an inductive process.

The inductive coding process for each dimension was carried out in two cycles, as suggested by Saldaña [76]. The first author (Coder 1) and last author (Coder 2) each coded half of the papers, cross-checking each other's work in an open coding approach. During this process, loose codes were developed to describe each of the dimensions, and new codes were introduced when they provided a more informative or distinctive categorization than the existing code set. In the second cycle, Coder 1 applied pattern coding to reduce the number of codes from the open coding cycle and to categorize these initial codes. Coder 2 reviewed the code set developed during the pattern coding cycle and discussed the existing code book as well as additions to the code set. Subsequently, simulator fidelity was assessed based on the categorization developed by Wynne et al. [94] and to reduce the various scenario codes into sub-dimensions DRIVING TASK, OCCURRENCE OF TRAFFIC, and REGULATION. Coder 1 then re-coded all papers with the final set of codes.

4 RESULTS

In the following, we present the results of our analysis of the 34 papers structured according to the dimensions described in Table 1. Figure 2 shows the number of reviewed papers by their publication year. A visualization of the results is shown in Figure 3. The models we reviewed can generally be divided into two types: (1) models that predict latent states (e.g., distraction, workload) based on physiological measures or on features that describe driver behavior, or (2) models that predict behavior (e.g., steering or task interleaving) based on design artifacts (e.g., UI layout, display size) or information derived from the driving environment (e.g., lane position).

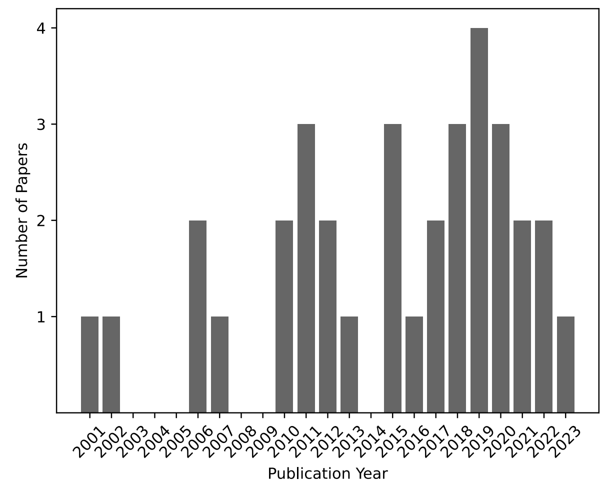

Figure 2: The number of reviewed papers with their publication year.

4.1 Target Variable – What to Predict?

The target variables predicted by the models can be categorized under the umbrella terms VISION, DRIVING, TASK, and MENTAL. Models that predict glance behavior (VISION) mostly predict metrics such as mean glance duration (4 papers), total glance duration (5), or number of glances (6). Other variables that are modeled include dwell times or occlusion-related metrics such as shutter open times. Models that predict DRIVING-related variables mostly focus on the prediction of lateral (8) and longitudinal (5) control but also on more abstract variables such as crash risk (1). The TASK category encompasses performance measurements related to NDRTs such as task completion time (8), task reaction time (1), or task switching probability (1). MENTAL target variables include the degree of distraction (4) or workload metrics (4). With regard to the latter, we found that the terms workload and cognitive load are not used consistently and are often confused. For example, it is not always clear whether authors are referring to cognitive load, with its origins in educational psychology [85], or to workload theory from the engineering and cognitive sciences [91]. Traditionally, however, papers in HCI are most often concerned with workload as measured by the NASA Task Load Index (NASA-TLX) [24]. Overall, our analysis shows that the majority of the papers target VISION-related variables (n = 16 (47%)), followed by DRIVING-related variables (n = 11 (32%)), TASK-related variables (n = 9 (26%)), and MENTAL variables (n = 8 (23%)). Only three models (9%) predict a combination of variables from different categories. They predict DRIVING together with VISION [58], MENTAL [2], or TASK [79].

Table 1: Eight modeling dimensions that were determined a-priori and reflect the topics we consider for further analysis.

Dimension	Description
TARGET VARIABLE	The specific aspect that the model predicts
MODEL INPUTS	The data the model uses for prediction
MODEL TYPE	The computational approach used
SYSTEM INTERACTION	The type of NDRT that is being modeled
MODALITY	Modality used to perform the NDRT
SCENARIO	Description of the driving scenario
ENVIRONMENT	The environment in which the data was collected or the model evaluated
INTENDED APPLICATION AREA	The use case for which the model was developed

4.2 Model Inputs – What Kind of Data is Used for Modeling?

The models reviewed in this paper predict the above-introduced variables using a variety of different features (i.e., model inputs). We identified the the following categories: UI DESIGN, DRIVING BEHAVIOR, GLANCE BEHAVIOR, PHYSIOLOGICAL MEASUREMENTS, SCENARIO INFORMATION and OTHER. UI DESIGN contains any variables that describe some aspect of the interface (e.g., display size (1) or button types (2)) and variables that are derived or manually specified based on the design of the interface (e.g., KLM operator sequences (4), SEEV values (2), or ACT-R production rules (3)). While the latter are not inputs in the traditional sense, they describe the effect that the design has on the interaction and thus on the prediction. 56% of all models consider features of this category. DRIVING BEHAVIOR contains variables that are connected to the driving task itself, e.g. lane position (6), vehicle speed (7) braking behavior (3), acceleration behavior (2). 32% of models use driving-related variables as input. GLANCE BEHAVIOR includes vision related measures such as eye movements (1), eyes-off-road-time (1) or blinking behavior (1). These metrics are often used to predict driver states such as distraction or workload. The same applies to PHYSIOLOGICAL MEASUREMENTS that are used in 9% of the papers and describe physiological measurements (e.g., heart rate or electrodermal activity). SCENARIO INFORMATION describes information such as road curvature (2) or target speed (2). OTHER is a catch-all category that covers a broad range of different inputs that did not fit well into the other categories, such as demand scores for different tasks obtained from the literature or acoustic data from the vehicle interior. Our analysis shows that most papers (n = 25 (74%)) use input variables from only one category. When categories are combined, the combinations often include DRIVING BEHAVIOR (n = 7 (77% of combinations)). Two out of three papers using PHYSIOLOGICAL MEASUREMENTS combine them with other variables [8, 27]. Three papers (33% of the combinations) combine GLANCE BEHAVIOR with other variables [8, 27, 53].

4.3 Model Types – ML, Cognitive Architectures, or Something Else?

This dimension describes the different categories of models used by the reviewed papers. Some papers compare models of different categories. In this case the paper is listed in both categories. We have identified the following types of models: ML, COGNITIVE ARCHITECTURES, REGRESSION, STOCHASTIC, GOMS, OTHER, MECHANISTIC, and QN. STOCHASTIC models are Hidden Markov Modelss (HMMs), Bayesian filters, or particle filters. Papers that use Fitts' law for modeling belong to MECHANISTIC. ACT-R, SEEV, or IMPRINT are examples for COGNITIVE ARCHITECTURES. GOMS includes all modeling approaches that use GOMS, but especially KLMs. Models that do not fit into any of these categories are listed under OTHER. Our analysis shows that most approaches rely on ML or COGNITIVE ARCHITECTURES (e.g., ACT-R, SEEV, or IMPRINT) (8, 24% each). Most of the papers in the ML category compare their models to other ML or REGRESSION modeling approaches. Jokinen et al. [41] presents the only approach that uses RL. Five papers (15%) use REGRESSION models for prediction, while 4 papers (12%) [51, 73, 74, 83] use GOMS and in particular KLMs. These models are used to predict task or glance times for interactions with in-vehicle interfaces. We also identified one paper that uses QN for modeling driver's NDRT behavior [57]. Four works are coded as OTHER [2, 25, 55, 98] (12%).

4.4 System Interaction – Which NDRTs are Modeled?

This dimension describes the interaction that is modeled. We identify the following categories: IVIS, MOBILE PHONE, CONVERSATION, and OTHER. Two-thirds (n = 23 (67%)) of approaches model IVIS interactions. These interactions can be divided into *real-world tasks* (13), *artificial tasks* (10), and *semi-artificial tasks* (1)[2]. *Real-world tasks* use consumer-grade in-vehicle interfaces or an interface that is similar in design and tasks that can be performed (e.g., reading a map, making a phone call). In contrast, artificial tasks serve as surrogate tasks, often manipulating only one specific feature (e.g., the number of elements on the screen) to increase controllability. *Semi-artificial tasks* are real-world tasks, such as making a phone call, but performed on a task-specific interface designed for the specific modeling task (c.f., [55]) MOBILE PHONE interactions are modeled in 10 papers (29%) and include interactions such as calling, texting or dialing. Approaches that model conversational interactions (CONVERSATION) make up 15% of all approaches. These conversational interactions include talking to a passenger or performing an auditory n-back task. OTHER describes the approach of Pekkanen et al. [72] who model NDRTs indirectly via an occlusion experiment. Our evaluation also shows that 20 papers (59%) present computational models that deal only with IVIS interaction and six papers (17%) deal only with MOBILE PHONE. Only 6 papers

[2] One study Large et al. [48] modeled both an artificial and a real task.

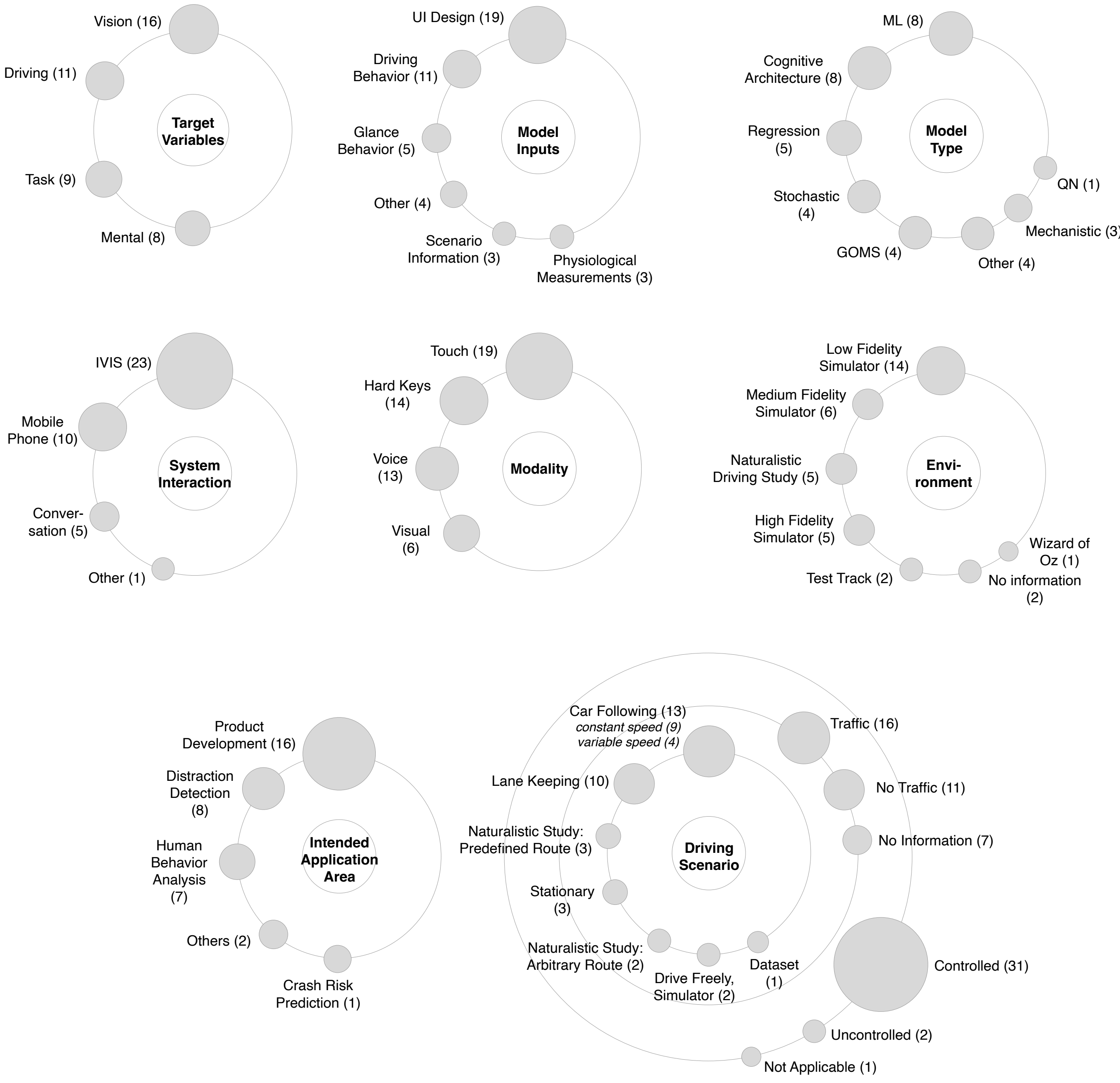

Figure 3: Orbit diagrams showing the coding dimension with the identified categories from the analysis.

(17%) combine system interactions from different categories. Li and Busso [53] investigate interactions from three categories, namely IVIS, MOBILE PHONE and CONVERSATION.

4.5 Modality – Touch or No Touch?

Our analysis shows the majority of works investigates TOUCH tasks ($n = 19\,(56\%)$), followed by HARD KEYS ($n = 14\,(41\%)$), VOICE ($n = 13\,(38\%)$) and VISUAL tasks. The latter category describes tasks where the interaction was either conversational (e.g., reading aloud visual information on the screen) or indirect (e.g., occlusion). For works that model interactions with smartphones or mobile phones but do not list the specific phone, we assume TOUCH for all papers that were published after 2013[3]. In total, 14 papers (41%) investigate tasks with modalities from more than one category. The most frequent combinations include VOICE, namely

[3]See the steep increase in smartphone sales during this time: https://www.statista.com/statistics/263437/global-smartphone-sales-to-end-users-since-2007/

VOICE + TOUCH (n = 6 (43% of all combinations)) and VOICE + HARD KEYS (n = 5 (35% of all combinations)). The papers by Li and Busso [53], Muñoz et al. [62] investigates different tasks from three modality categories, namely TOUCH, HARD KEYS, and VOICE.

4.6 Environment – Simulator Studies vs. Naturalistic Driving Studies

According to the framework of Janssen et al. [36], the environment can be described as *"the world being modeled"*. Accordingly, environments can range from *simulated* as in a driving simulator to *real* as in naturalistic driving study. Our analysis reveals that the majority (n = 26 (76%)) of the works use driving simulators to evaluate their models or to gather data for their models. Since the fidelity of a driving simulator can influence driver behavior [29], we further categorize each driving simulator according to the classification theme proposed by Wynne et al. [94]. Most of the simulators are LOW-FIDELITY (n = 14 (53% of all simulators)), followed by MEDIUM-FIDELITY (23%) and HIGH-FIDELITY (19%). For real-world driving studies, we distinguish between TEST-TRACK and NATURALISTIC DRIVING STUDIES. While NATURALISTIC DRIVING STUDIES were used in five of the papers (14%), TEST-TRACK was used two times in total. A WIZARD-OF-OZ simulator was used in only one paper [82]. The paper by Pekkanen et al. [72] is the only one that compares driver performance in two different environments, a MEDIUM-FIDELITY simulator and on a TEST-TRACK.

4.7 Scenario – Controlled or Uncontrolled?

Janssen et al. [36] describe scenarios *"the way one moves through the world"*, denoting *"what types of situations are encountered or not"*. To describe the scenarios we introduce the following sub-dimension: DRIVING TASK, OCCURRENCE OF TRAFFIC and REGULATION. DRIVING TASK refers to scenarios, such as LANE KEEPING or CAR FOLLOWING, which are mostly used in driving simulators. For naturalistic studies, this sub-dimension includes whether the participants drove a predefined route or were allowed to drive wherever they wanted. TRAFFIC describes the occurrence of other traffic participants. REGULATION describes the degree to which the scenarios have been controlled. In a fully CONTROLLED scenario, each participant experiences the exact same task, route, and traffic conditions at the exact same time. In an UNCONTROLLED scenario, on the other hand, drivers may be able to choose the route they take and the task they perform. An example of the latter would be data collected from different customers across the country. Our analysis shows that the majority of studies were conducted in a CONTROLLED setting (n = 31 (91%)). Only Ebel et al. [18] and Young [97] use data from an UNCONTROLLED scenario. The ratio of papers with TRAFFIC (n = 16 (47%)) and WITHOUT TRAFFIC (n = 11 (33%)) is almost equal. However, seven papers (20%) do not report whether their scenario included traffic. Regarding the DRIVING TASK, the papers using a simulator are almost equally divided between CAR FOLLOWING scenarios (n = 13 (50% of the simulator studies)) and LANE KEEPING scenarios (n = 10 (38%)). An exception are the works by Watkins et al. [88] and Chen et al. [8] who allowed their participants to drive relatively freely in a simulator. Of the NATURALISTIC DRIVING STUDIES, most used a predefined route (n = 3 (60% of naturalistic studies)).

4.8 Intended Application Area - What's the Goal of the Model?

This dimension describes the application the models are targeted at and where the authors anticipate these models to be valuable. Our analysis shows that nearly the half of the developed models aim to be valuable in the *product development* of in-vehicle interfaces (n = 16 (47%)). The second most targeted application is DISTRACTION DETECTION (n = 8 (24%)) followed by *human behavior analysis* (20%). The latter includes models that aim to *understand human behavior and improve performance* [49], *create plausible models for car following* [72], or predict *interleaving strategies in multitasking* [37, 38, 41, 46]. The model proposed by Young [97] is intended for CRASH RISK PREDICTION.

5 RESEARCH GAPS

Our literature review highlights several research gaps, which we discuss below. They all contribute directly or indirectly to the development of more accurate computational models for the design and evaluation of in-vehicle UIs.

Tools for application of models to HCI problems. Of all 34 models that we reviewed, only three [25, 79, 83] are associated with a tool that allows users to operationalize the predictions made by the model. The tool presented by Harvey and Stanton [25] is a spreadsheet that allows users to specify sequences of manual, visual, and cognitive operators that serve as an input to the critical path analysis. Schneegaß et al. [83] present a tool that, lets users annotate UI screenshots with UI elements to then automatically compute a KLM. The most advanced tool is presented by Salvucci et al. [81]. Distract-R is based on an ACT-R model and allows users to prototype basic interfaces from which the model predicts driving-related metrics. These metrics serve as proxies for evaluating the distraction potential of early stage UI designs. In addition to the fact that only 3 out of 34 models were integrated into tools it needs to be emphasized that these tools are between 10 - 18 years old. Therefore downstream research should look into operationalizing state of the art computational models into accessible tools that allows leveraging their benefits.

Scenarios are simplistic and model predictions context specific. Almost all models (except of the work presented by Ebel et al. [18]) are developed and evaluated based on small data sets collected in controlled (simulator) environments with simplistic driving scenarios (e.g., lane keeping or car following tasks). Furthermore 18 out of 34 scenarios do not consider other traffic. While these models may make accurate predictions in the context in which they were developed, it is not clear how they generalize to real-world problems with complex *and* dynamic driving scenarios. Thus, future research and future models should address what is known in robotics as the Sim2Real gap [100] by including more diverse and realistic scenarios, which helps to increase traffic safety through models that generalize to real-world conditions.

Large open datasets. While there are large publicly available datasets for driver distraction detection (e.g., [60]), there are no such datasets that also provide detailed information about driver interaction with in-vehicle interfaces. Publicly available datasets from naturalistic driving studies [13, 86] do not contain user interaction or interface design information and are therefore not useful for the modeling approaches reviewed in this work. Accordingly, the few models that are based on large datasets from real world driving or naturalistic driving studies [18] are developed in cooperation with Original Equipment Manufacturers (OEMs). In this context, it is interesting to note that while OEMs are interested in such models and see their potential [19, 64], they do not conduct or publish much research in this direction. We argue that future naturalistic driving studies should also collect detailed interaction and behavioral data. Also, given that OEMs are already lagging behind digital companies when it comes to data-driven UX evaluation [19], they should consider being more open so that they can leverage the progress being made in the research community.

Grand modeling challenges. In total we count 21 different target variables that the models in this review predict. Furthermore the datasets used in this approaches are almost always specific to the respective model. While this is not bad in itself, it shows that there is no consensus toward grand challenges for designing models of future interaction behavior with in-vehicle interfaces. While recognizing the various different facets of driver modeling, we argue that grand challenges for modeling, such as those established in other domains (e.g., DARPA challenge [6], ImageNet [45]), could benefit progress toward better and more accurate models.

Models that generalize to arbitrary interface designs. Most of the models in this review are very specific to the exact use case they were developed for and require manual specification as soon as the task or design to be evaluated changes. This applies for example to models that rely on ACT-R [46, 79, 80]. While they might provide accurate and transparent predictions, the production rules must be manually specified by experts. The same applies to models based on the KLM technique [51, 73, 74, 83] or SEEV [31]. A valuable next step would be to develop models that generalize to arbitrary interface designs. This can be achieved, for example, by models trained on generic data such as pixels. In this way, any screenshot or driving scene generated by a simulation could be used as input to the model.

Consideration of interpersonal and demographic factors. Of the models in this review only one approach [51] incorporates interpersonal differences or demographic factors for prediction. However, related work shows that age, cultural background, and other individual characteristics impact driving behavior [30, 95], distraction [65] or the willingness to engage in NDRTs [23]. Leveraging this approach could enhance the modeling of driver behavior and improve the design of in-vehicle interfaces. Outside of the automotive domain, approaches such as data-driven personas, incorporating demographic information, are already successfully used [77, 99].

6 DISCUSSION

Most of the models reviewed in this paper aim to improve the design or evaluation of in-vehicle interfaces. Still, there is no documentation of such models being used in the industrial UX design and evaluation process. Our results presented in section 4 and research gaps identified in section 5 elucidate several open research questions. Drawing connections with related work, we formulate research potentials for future work towards computational models that can support the automotive UI design process.

Align computational modeling with real-world design problems. Our analysis shows that most models seem to focus on explaining isolated psychological or behavioral phenomena (e.g. workload [42, 47, 82], task reaction time [55] or the scenario in which the task is performed [88]) by replicating experimental results from controlled user studies, rather than solving concrete HCI problems. However, to improve the way in-vehicle interfaces are designed today, we propose a shift towards computational models that directly target the specific problems within the automotive UI design process. Similar to what Janssen et al. [34] call *"[f]rom description to prescription"* [34, p. 3], we argue that computational models for in-vehicle user interfaces must be aligned with real-world HCI problems and oriented toward the needs that automotive designers and safety engineers face. Looking at the needs of HCI professionals in the automotive domain, Ebel et al. [16] argue that practitioners need models and tools that can be integrated into their design process and that allow early design evaluation for metrics such as distraction potential and usability. This would imply that future models (1) must be able to make predictions based on features derived directly from prototypical UI designs and mock-ups (c.f., [18]), (2) models must be integrated into tools that allow non-experts to use them (c.f., [81, 83]), and (3) predictions must be accurate enough to drive design decisions in such a high-stakes (economic and safety) environment.

Move from models that need manual specification to models that adapt to new data and new problems. Our analysis shows that (1) there is a gap for models that are not only generalizable to different interface designs, but also generalizable to different scenarios and that (2) many models rely on manual specifications performed by experts [74, 79, 93]. To move from models that need manual specification to models that adapt to new data and new problems wee see two research potentials: (1) data-driven ML-based models that are trained on large general data (see section 5) that is continuously updated, allowing the models to generalize over different designs and adapt to distribution shifts in the data, or (2) computationally rational models [67] whose internal information processing mechanisms are grounded in theory and whose behavioral policies are approximated via RL (e.g., [41]). The latter combination can be a promising solution to the trade-off between explainability and predictive accuracy often faced by supervised ML-based approaches [10].

Increase openness and transparency. An analysis of where the final set of papers was published shows that researchers from a variety of backgrounds (psychology, human factors, engineering, HCI) are developing computational models of driver interaction behavior. This diversity in backgrounds makes it particular important to be transparent and open. However, our review discloses a lack of shared source code and data (only 3 papers shared artifacts [41, 51, 72]), ambiguous model descriptions, and inconsistent terminology, all of which hinder replicability, reproducability and eventually progress. We argue that more openness in terms of artifact sharing and more transparency in reporting would benefit the model development in general. For the latter, a modeling taxonomy that unifies terminology can foster collaboration and bridge gaps between psychology, human factors, engineering, and HCI. The dimensions developed in this paper can serve as a basis for such a taxonomy.

Limitations

As with any literature review, we acknowledge the potential for coding errors or subjective interpretation. However, because each paper was independently reviewed by two authors and disagreements were resolved in discussion, these errors should be minimized. Furthermore, the ambiguity in the model descriptions, the large number of different phenomena being modeled and the variety research field the models come from led to many discussions among the coders to define appropriate categories. We want to state that this categorization is *one* categorization that can be used, but may not be the best. However, we do not claim to have developed a taxonomy and see this as future work (see section 5).

ACKNOWLEDGMENTS

We thank Christian P Janssen for the initial discussions that led to our research questions, and everyone involved in the Cadima project for making such a tool freely available.

SUPPLEMENTARY MATERIAL

All research artifacts (codebook, search strings, references after title - abstract screening) and the final coding results can be found at https://osf.io/qgxty/

REFERENCES

[1] National Highway Traffic Safety Administration and others. 2012. Visual-manual NHTSA driver distraction guidelines for in-vehicle electronic devices. *Washington, DC: National Highway Traffic Safety Administration (NHTSA), Department of Transportation (DOT)* (2012).

[2] Andrei Aksjonov, Pavel Nedoma, Valery Vodovozov, Eduard Petlenkov, and Martin Herrmann. 2019. Detection and Evaluation of Driver Distraction Using Machine Learning and Fuzzy Logic. *IEEE Transactions on Intelligent Transportation Systems* 20, 6 (June 2019), 2048–2059. https://doi.org/10.1109/TITS.2018.2857222

[3] John R. Anderson, Michael Matessa, and Christian Lebiere. 1997. ACT-R: A Theory of Higher Level Cognition and Its Relation to Visual Attention. *Human–Computer Interaction* 12, 4 (Dec. 1997), 439–462. https://doi.org/10.1207/s15327051hci1204_5

[4] Matthew Avery. 2024. Bring Back Buttons. https://www.linkedin.com/feed/update/urn:li:activity:7170356955910254592/

[5] Gilles Bailly, Antti Oulasvirta, Timo Kötzing, and Sabrina Hoppe. 2013. MenuOptimizer: interactive optimization of menu systems. In *Proceedings of the 26th annual ACM symposium on User interface software and technology*. ACM, St. Andrews Scotland, United Kingdom, 331–342. https://doi.org/10.1145/2501988.2502024

[6] Martin Buehler, Karl Iagnemma, and Sanjiv Singh. 2007. *The 2005 DARPA grand challenge: the great robot race*. Vol. 36. springer.

[7] Stuart Card. 1983. *The psychology of human-computer interaction.* L. Erlbaum Associates, Hillsdale, N.J.

[8] Jie Chen, YaNan Jiang, ZhiXiang Huang, XiaoHui Guo, BoCai Wu, Long Sun, and Tao Wu. 2021. Fine-Grained Detection of Driver Distraction Based on Neural Architecture Search. *IEEE Transactions on Intelligent Transportation Systems* 22, 9 (Sept. 2021), 5783–5801. https://doi.org/10.1109/TITS.2021.3055545

[9] Xiuli Chen, Gilles Bailly, Duncan P. Brumby, Antti Oulasvirta, and Andrew Howes. 2015. The Emergence of Interactive Behavior: A Model of Rational Menu Search. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*. ACM, Seoul Republic of Korea, 4217–4226. https://doi.org/10.1145/2702123.2702483

[10] Barnaby Crook, Maximilian Schlüter, and Timo Speith. 2023. Revisiting the Performance-Explainability Trade-Off in Explainable Artificial Intelligence (XAI). In *2023 IEEE 31st International Requirements Engineering Conference Workshops (REW)*. IEEE, Hannover, Germany, 316–324. https://doi.org/10.1109/REW57809.2023.00060

[11] Biplab Deka, Zifeng Huang, Chad Franzen, Joshua Hibschman, Daniel Afergan, Yang Li, Jeffrey Nichols, and Ranjitha Kumar. 2017. 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, Québec City QC Canada, 845–854. https://doi.org/10.1145/3126594.3126651

[12] Thomas A. Dingus, Feng Guo, Suzie Lee, Jonathan F. Antin, Miguel Perez, Mindy Buchanan-King, and Jonathan Hankey. 2016. Driver crash risk factors and prevalence evaluation using naturalistic driving data. *Proceedings of the National Academy of Sciences* 113, 10 (March 2016), 2636–2641. https://doi.org/10.1073/pnas.1513271113

[13] T. A. Dingus, S. G. Klauer, V. L. Neale, A. Petersen, S. E. Lee, J. Sudweeks, M. A. Perez, J. Hankey, D. Ramsey, S. Gupta, C. Bucher, Z. R. Doerzaph, J. Jermeland, and R. R. Knipling. 2006. The 100-Car Naturalistic Driving Study: Phase II - Results of the 100-Car Field Experiment. https://doi.org/10.1037/e624282011-001

[14] Anup Doshi and Mohan M. Trivedi. 2011. Tactical driver behavior prediction and intent inference: A review. In *2011 14th International IEEE Conference on Intelligent Transportation Systems (ITSC)*. IEEE, Washington, DC, USA, 1892–1897. https://doi.org/10.1109/ITSC.2011.6083128

[15] Patrick Ebel, Moritz Berger, Christoph Lingenfelder, and Andreas Vogelsang. 2022. 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*. ACM, Seoul Republic of Korea, 263–273. https://doi.org/10.1145/3543174.3545173

[16] Patrick Ebel, Florian Brokhausen, and Andreas Vogelsang. 2020. The Role and Potentials of Field User Interaction Data in the Automotive UX Development Lifecycle: An Industry Perspective. In *12th International Conference on Automotive User Interfaces and Interactive Vehicular Applications*. ACM, Virtual Event DC USA, 141–150. https://doi.org/10.1145/3409120.3410638

[17] Patrick Ebel, Christoph Lingenfelder, and Andreas Vogelsang. 2023. Multitasking While Driving: How Drivers Self-Regulate Their Interaction with In-Vehicle Touchscreens in Automated Driving. *International Journal of Human–Computer Interaction* (June 2023), 1–18. https://doi.org/10.1080/10447318.2023.2215634

[18] Patrick Ebel, Christoph Lingenfelder, and Andreas Vogelsang. 2023. On the forces of driver distraction: Explainable predictions for the visual demand of in-vehicle touchscreen interactions. *Accident Analysis & Prevention* 183 (April 2023), 106956. https://doi.org/10.1016/j.aap.2023.106956

[19] Patrick Ebel, Julia Orlovska, Sebastian Hünemeyer, Casper Wickman, Andreas Vogelsang, and Rikard Söderberg. 2021. Automotive UX design and data-driven development: Narrowing the gap to support practitioners. *Transportation Research Interdisciplinary Perspectives* 11 (Sept. 2021), 100455. https://doi.org/10.1016/j.trip.2021.100455

[20] Hilkka Grahn and Tuomo Kujala. 2020. 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* 142 (Oct. 2020), 102467. https://doi.org/10.1016/j.ijhcs.2020.102467

[21] Andrew Gross. 2017. New Vehicle Infotainment Systems Create Increased Distractions Behind the Wheel. https://newsroom.aaa.com/2017/10/new-vehicle-infotainment-systems-create-increased-distractions-behind-wheel/

[22] Andrew Gross. 2019. In-Vehicle Infotainment Systems Especially Distracting to Older Drivers. https://newsroom.aaa.com/2019/07/vehicle-infotainment-systems-especially-distracting-older-drivers/

[23] Feng Guo, Sheila G Klauer, Youjia Fang, Jonathan M Hankey, Jonathan F Antin, Miguel A Perez, Suzanne E Lee, and Thomas A Dingus. 2016. The effects of age on crash risk associated with driver distraction. *International Journal of Epidemiology* (Oct. 2016), dyw234. https://doi.org/10.1093/ije/dyw234

[24] Sandra G Hart. 1986. NASA task load index (TLX). (1986). https://ntrs.nasa.gov/api/citations/20000021487/downloads/20000021487.pdf

[25] Catherine Harvey and Neville A. Stanton. 2013. Modelling the hare and the tortoise: predicting the range of in-vehicle task times using critical path analysis. *Ergonomics* 56, 1 (Jan. 2013), 16–33. https://doi.org/10.1080/00140139.2012.733031

[26] Catherine Harvey, Neville A. Stanton, Carl A. Pickering, Mike McDonald, and Pengjun Zheng. 2011. In-Vehicle Information Systems to Meet the Needs of Drivers. *International Journal of Human-Computer Interaction* 27, 6 (June 2011), 505–522. https://doi.org/10.1080/10447318.2011.555296

[27] Dengbo He, Ziquan Wang, Elias B. Khalil, Birsen Donmez, Guangkai Qiao, and Shekhar Kumar. 2022. Classification of Driver Cognitive Load: Exploring the Benefits of Fusing Eye-Tracking and Physiological Measures. *Transportation Research Record: Journal of the Transportation Research Board* 2676, 10 (Oct. 2022), 670–681. https://doi.org/10.1177/03611981221090937

[28] Ahmadreza Mahmoudzadeh Hesamoddin Razi-Ardakani and Mohammad Kermanshah. 2019. What factors results in having a severe crash? a closer look on distraction-related factors. *Cogent Engineering* 6, 1 (2019), 1708652. https://doi.org/10.1080/23311916.2019.1708652 Publisher: Cogent OA tex.eprint: https://doi.org/10.1080/23311916.2019.1708652.

[29] Chantal Himmels, Joost Venrooij, Arben Parduzi, Martin Peller, and Andreas Riener. 2024. The bigger the better? Investigating the effects of driving simulator fidelity on driving behavior and perception. *Transportation Research Part F: Traffic Psychology and Behaviour* 101 (Feb. 2024), 250–266. https://doi.org/10.1016/j.trf.2024.01.007

[30] Tim Horberry, Janet Anderson, Michael A. Regan, Thomas J. Triggs, and John Brown. 2006. Driver distraction: The effects of concurrent in-vehicle tasks, road environment complexity and age on driving performance. *Accident Analysis & Prevention* 38, 1 (Jan. 2006), 185–191. https://doi.org/10.1016/j.aap.2005.09.007

[31] William J. Horrey, Christopher D. Wickens, and Kyle P. Consalus. 2006. Modeling drivers' visual attention allocation while interacting with in-vehicle technologies. *Journal of Experimental Psychology: Applied* 12, 2 (2006), 67–78. https://doi.org/10.1037/1076-898X.12.2.67

[32] Karel Hurts, Linda S. Angell, and Miguel A. Perez. 2011. The Distracted Driver: Mechanisms, Models, and Measurement. *Reviews of Human Factors and Ergonomics* 7, 1 (Sept. 2011), 3–57. https://doi.org/10.1177/1557234X11410387

[33] Aleksi Ikkala, Florian Fischer, Markus Klar, Miroslav Bachinski, Arthur Fleig, Andrew Howes, Perttu Hämäläinen, Jörg Müller, Roderick Murray-Smith, and Antti Oulasvirta. 2022. Breathing Life Into Biomechanical User Models. In *Proceedings of the 35th Annual ACM Symposium on User Interface Software and Technology*. ACM, Bend OR USA, 1–14. https://doi.org/10.1145/3526113.3545689

[34] Christian P. Janssen, Martin Baumann, and Antti Oulasvirta. 2024. Computational models of cognition for human-automated vehicle interaction: State-of-the-art and future directions. *International Journal of Human-Computer Studies* 185 (May 2024), 103230. https://doi.org/10.1016/j.ijhcs.2024.103230

[35] Christian P. Janssen, Martin Baumann, Antti Oulasvirta, Shamsi Tamara Iqbal, and Luisa Heinrich. 2022. *Computational Models of Human-Automated Vehicle Interaction (Dagstuhl Seminar 22102)*. Technical Report. [object Object]. 67 pages, 7782081 bytes pages. https://doi.org/10.4230/DAGREP.12.3.15 Artwork Size: 67 pages, 7782081 bytes ISSN: 2192-5283 Medium: application/pdf.

[36] Christian P. Janssen, Linda Ng Boyle, Wendy Ju, Andreas Riener, and Ignacio Alvarez. 2020. Agents, environments, scenarios: A framework for examining models and simulations of human-vehicle interaction. *Transportation Research Interdisciplinary Perspectives* 8 (2020), 100214. https://doi.org/10.1016/j.trip.2020.100214

[37] Christian P. Janssen and Duncan P. Brumby. 2010. Strategic Adaptation to Performance Objectives in a Dual-Task Setting. *Cognitive Science* 34, 8 (Nov. 2010), 1548–1560. https://doi.org/10.1111/j.1551-6709.2010.01124.x

[38] Christian P. Janssen, Duncan P. Brumby, and Rae Garnett. 2012. Natural Break Points: The Influence of Priorities and Cognitive and Motor Cues on Dual-Task Interleaving. *Journal of Cognitive Engineering and Decision Making* 6, 1 (March 2012), 5–29. https://doi.org/10.1177/1555343411432339

[39] Bonnie John. 1995. Why goms? *interactions* 2, 4 (1995), 80–89. Publisher: ACM New York, NY, USA.

[40] Jussi Jokinen, Aditya Acharya, Mohammad Uzair, Xinhui Jiang, and Antti Oulasvirta. 2021. Touchscreen Typing As Optimal Supervisory Control. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. ACM, Yokohama Japan, 1–14. https://doi.org/10.1145/3411764.3445483

[41] Jussi P. P. Jokinen, Tuomo Kujala, and Antti Oulasvirta. 2020. Multitasking in Driving as Optimal Adaptation Under Uncertainty. *Human Factors: The Journal of the Human Factors and Ergonomics Society* 63, 8 (July 2020), 1324–1341. https://doi.org/10.1177/0018720820927687 Publisher: SAGE Publications.

[42] Cansu Kandemir, Holly A.H. Handley, and Deborah Thompson. 2018. A workload model to evaluate distracters and driver's aids. *International Journal of Industrial Ergonomics* 63 (Jan. 2018), 18–36. https://doi.org/10.1016/j.ergon.2016.09.004

[43] Barbara Kitchenham, Stuart Charters, and others. 2007. Guidelines for performing systematic literature reviews in software engineering.

[44] Christian Kohl, Emma J. McIntosh, Stefan Unger, Neal R. Haddaway, Steffen Kecke, Joachim Schiemann, and Ralf Wilhelm. 2018. Online tools supporting the conduct and reporting of systematic reviews and systematic maps: a case study on CADIMA and review of existing tools. *Environmental Evidence* 7, 1 (Feb. 2018). https://doi.org/10.1186/s13750-018-0115-5 Publisher: Springer Science and Business Media LLC.

[45] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E. Hinton. 2017. ImageNet classification with deep convolutional neural networks. *Commun. ACM* 60, 6 (May 2017), 84–90. https://doi.org/10.1145/3065386

[46] Tuomo Kujala and Dario D. Salvucci. 2015. Modeling visual sampling on in-car displays: The challenge of predicting safety-critical lapses of control. *International Journal of Human-Computer Studies* 79 (July 2015), 66–78. https://doi.org/10.1016/j.ijhcs.2015.02.009

[47] Shekhar Kumar, Dengbo He, Guangkai Qiao, and Birsen Donmez. 2022. Classification of Driver Cognitive Load based on Physiological Data: Exploring Recurrent Neural Networks. In *2022 International Conference on Advanced Robotics and Mechatronics (ICARM)*. IEEE, Guilin, China, 19–24. https://doi.org/10.1109/ICARM54641.2022.9959588

[48] David R. Large, Gary Burnett, Elizabeth Crundall, Editha Van Loon, Ayse L. Eren, and Lee Skrypchuk. 2018. Developing Predictive Equations to Model the Visual Demand of In-Vehicle Touchscreen HMIs. *International Journal of Human–Computer Interaction* 34, 1 (Jan. 2018), 1–14. https://doi.org/10.1080/10447318.2017.1306940

[49] David R. Large, Elizabeth Crundall, Gary Burnett, and Lee Skrypchuk. 2015. Predicting the visual demand of finger-touch pointing tasks in a driving context. In *Proceedings of the 7th International Conference on Automotive User Interfaces and Interactive Vehicular Applications*. ACM, Nottingham United Kingdom, 221–224. https://doi.org/10.1145/2799250.2799256

[50] John Lee, Kristie Young, and Michael Regan. 2008. Defining Driver Distraction. In *Driver Distraction*, Michael Regan, John Lee, and Kristie Young (Eds.). CRC Press, 31–40. https://doi.org/10.1201/9781420007497.ch3

[51] Seul Chan Lee, Sol Hee Yoon, and Yong Gu Ji. 2019. Modeling task completion time of in-vehicle information systems while driving with keystroke level modeling. *International Journal of Industrial Ergonomics* 72 (July 2019), 252–260. https://doi.org/10.1016/j.ergon.2019.06.001

[52] Luis A. Leiva, Morteza Shiripour, and Antti Oulasvirta. 2023. Modeling how different user groups perceive webpage aesthetics. *Universal Access in the Information Society* 22, 4 (Nov. 2023), 1417–1424. https://doi.org/10.1007/s10209-022-00910-x

[53] Nanxiang Li and Carlos Busso. 2015. Predicting Perceived Visual and Cognitive Distractions of Drivers With Multimodal Features. *IEEE Transactions on Intelligent Transportation Systems* 16, 1 (Feb. 2015), 51–65. https://doi.org/10.1109/TITS.2014.2324414

[54] Yang Li, Samy Bengio, and Gilles Bailly. 2018. Predicting Human Performance in Vertical Menu Selection Using Deep Learning. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. ACM, Montreal QC Canada, 1–7. https://doi.org/10.1145/3173574.3173603

[55] Bin Lin and Changxu Wu. 2011. Mathematical modeling of the human cognitive system in two serial processing stages with its applications in adaptive workload-management systems. *IEEE Transactions on Intelligent Transportation Systems* 12, 1 (2011), 221–231. https://doi.org/10.1109/TITS.2010.2081359

[56] Yili Liu. 1996. Queueing network modeling of elementary mental processes. *Psychological Review* 103, 1 (1996), 116–136. https://doi.org/10.1037/0033-295X.103.1.116

[57] Yili Liu, Robert Feyen, and Omer Tsimhoni. 2006. Queueing Network-Model Human Processor (QN-MHP): A computational architecture for multitask performance in human-machine systems. *ACM Transactions on Computer-Human Interaction* 13, 1 (March 2006), 37–70. https://doi.org/10.1145/1143518.1143520

[58] Jun Ma, Zaiyan Gong, Jianjie Tan, Qianwen Zhang, and Yuanyang Zuo. 2020. Assessing the driving distraction effect of vehicle HMI displays using data mining techniques. *Transportation Research Part F: Traffic Psychology and Behaviour* 69 (Feb. 2020), 235–250. https://doi.org/10.1016/j.trf.2020.01.016

[59] Gerrit Meixner and Christian Müller (Eds.). 2017. *Automotive User Interfaces: Creating Interactive Experiences in the Car*. Springer International Publishing, Cham. https://doi.org/10.1007/978-3-319-49448-7

[60] Anna Montoya, Dan Holman, Taylor Smith, and Wendy Kan. 2016. State farm distracted driver detection. https://kaggle.com/competitions/state-farm-distracted-driver-detection

[61] Hee-Seung Moon, Antti Oulasvirta, and Byungjoo Lee. 2023. Amortized Inference with User Simulations. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*. ACM, Hamburg Germany, 1–20. https://doi.org/10.1145/3544548.3581439

[62] Mauricio Muñoz, Bryan Reimer, Joonbum Lee, Bruce Mehler, and Lex Fridman. 2016. Distinguishing patterns in drivers' visual attention allocation using Hidden Markov Models. *Transportation Research Part F: Traffic Psychology and Behaviour* 43 (Nov. 2016), 90–103. https://doi.org/10.1016/j.trf.2016.09.015

[63] Paul Myles. 2020. Distraction Study Calls Out Touchscreens. https://www.wardsauto.com/industry/distraction-study-calls-out-touchscreens

[64] Julia Orlovska, Casper Wickman, and Rikard Söderberg. 2018. Big Data Usage Can Be a Solution for User Behavior Evaluation: An Automotive Industry Example. *Procedia CIRP* 72 (2018), 117–122. https://doi.org/10.1016/j.procir.2018.03.102

[65] C. Ortiz, S. Ortiz-Peregrina, J.J. Castro, M. Casares-López, and C. Salas. 2018. Driver distraction by smartphone use (WhatsApp) in different age groups. *Accident Analysis & Prevention* 117 (Aug. 2018), 239–249. https://doi.org/10.1016/j.aap.2018.04.018

[66] Antti Oulasvirta, Niraj Ramesh Dayama, Morteza Shiripour, Maximilian John, and Andreas Karrenbauer. 2020. Combinatorial Optimization of Graphical User Interface Designs. *Proc. IEEE* 108, 3 (March 2020), 434–464. https://doi.org/10.1109/JPROC.2020.2969687

[67] Antti Oulasvirta, Jussi P. P. Jokinen, and Andrew Howes. 2022. Computational Rationality as a Theory of Interaction. In *CHI Conference on Human Factors in Computing Systems*. ACM, New Orleans LA USA, 1–14. https://doi.org/10.1145/3491102.3517739

[68] Antti Oulasvirta, Per Ola Kristensson, Xiaojun Bi, and Andrew Howes (Eds.). 2018. *Computational Interaction*. Vol. 1. Oxford University Press. https://doi.org/10.1093/oso/9780198799603.001.0001

[69] Matthew J. Page, Joanne E. McKenzie, Patrick M. Bossuyt, Isabelle Boutron, Tammy C. Hoffmann, Cynthia D. Mulrow, Larissa Shamseer, Jennifer M. Tetzlaff, Elie A. Akl, Sue E. Brennan, Roger Chou, Julie Glanville, Jeremy M. Grimshaw, Asbjørn Hróbjartsson, Manoj M. Lalu, Tianjing Li, Elizabeth W. Loder, Evan Mayo-Wilson, Steve McDonald, Luke A. McGuinness, Lesley A. Stewart, James Thomas, Andrea C. Tricco, Vivian A. Welch, Penny Whiting, and David Moher. 2021. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *Systematic Reviews* 10, 1 (March 2021). https://doi.org/10.1186/s13643-021-01626-4 Publisher: Springer Science and Business Media LLC.

[70] Sanna M. Pampel, Gary Burnett, Chrisminder Hare, Harpreet Singh, Arber Shabani, Lee Skrypchuk, and Alex Mouzakitis. 2019. 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 (AutomotiveUI '19)*. Association for Computing Machinery, New York, NY, USA, 254–261. https://doi.org/10.1145/3342197.3344538 Number of pages: 8 Place: Utrecht, Netherlands.

[71] Junho Park and Maryam Zahabi. 2022. A review of human performance models for prediction of driver behavior and interactions with in-vehicle technology. *Human Factors* 0, 0 (2022), 00187208221132740. https://doi.org/10.1177/00187208221132740 tex.eprint: https://doi.org/10.1177/00187208221132740.

[72] Jami Pekkanen, Otto Lappi, Paavo Rinkkala, Samuel Tuhkanen, Roosa Frantsi, and Heikki Summala. 2018. A computational model for driver's cognitive state, visual perception and intermittent attention in a distracted car following task. *Royal Society Open Science* 5, 9 (Sept. 2018), 180194. https://doi.org/10.1098/rsos.180194

[73] Michael Pettitt, Gary Burnett, and Alan Stevens. 2007. 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 (CHI '07)*. Association for Computing Machinery, New York, NY, USA, 1515–1524. https://doi.org/10.1145/1240624.1240852

[74] Christian Purucker, Frederik Naujoks, Andy Prill, and Alexandra Neukum. 2017. Evaluating distraction of in-vehicle information systems while driving by predicting total eyes-off-road times with keystroke level modeling. *Applied Ergonomics* 58 (Jan. 2017), 543–554. https://doi.org/10.1016/j.apergo.2016.04.012

[75] Christopher Rawles, Alice Li, Daniel Rodriguez, Oriana Riva, and Timothy Lillicrap. 2023. Android In The Wild: A large-scale dataset for android device control. In *Advances in neural information processing systems*, A. Oh, T. Neumann, A. Globerson, K. Saenko, M. Hardt, and S. Levine (Eds.), Vol. 36. Curran Associates, Inc., 59708–59728. https://proceedings.neurips.cc/paper_files/paper/2023/file/bbbb6308b402fe909c39dd29950c32e0-Paper-Datasets_and_Benchmarks.pdf

[76] Johnny Saldaña. 2013. *The coding manual for qualitative researchers* (2nd ed ed.). SAGE, Los Angeles. OCLC: ocn796279115.

[77] Joni Salminen, Kathleen Guan, Soon-Gyo Jung, and Bernard J. Jansen. 2021. A Survey of 15 Years of Data-Driven Persona Development. *International Journal of Human–Computer Interaction* 37, 18 (Nov. 2021), 1685–1708. https://doi.org/10.1080/10447318.2021.1908670

[78] Paul M. Salmon, Neville A. Stanton, and Kristie Lee Young. 2012. Situation awareness on the road: review, theoretical and methodological issues, and future directions. *Theoretical Issues in Ergonomics Science* 13, 4 (July 2012), 472–492. https://doi.org/10.1080/1463922X.2010.539289

[79] Dario D. Salvucci. 2001. Predicting the effects of in-car interface use on driver performance: an integrated model approach. *International Journal of Human-Computer Studies* 55, 1 (July 2001), 85–107. https://doi.org/10.1006/ijhc.2001.0472

[80] Dario D Salvucci and Kristen L Macuga. 2002. Predicting the effects of cellular-phone dialing on driver performance. *Cognitive Systems Research* 3, 1 (March 2002), 95–102. https://doi.org/10.1016/S1389-0417(01)00048-1

[81] Dario D. Salvucci, Mark Zuber, Ekaterina Beregovaia, and Daniel Markley. 2005. Distract-R: rapid prototyping and evaluation of in-vehicle interfaces. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, Portland Oregon USA, 581–589. https://doi.org/10.1145/1054972.1055052

[82] Maria Schmidt, Ojashree Bhandare, Ajinkya Prabhune, Wolfgang Minker, and Steffen Werner. 2020. Classifying Cognitive Load for a Proactive In-car Voice Assistant. In *2020 IEEE Sixth International Conference on Big Data Computing Service and Applications (BigDataService)*. IEEE, Oxford, United Kingdom, 9–16. https://doi.org/10.1109/BigDataService49289.2020.00010

[83] Stefan Schneegaß, Bastian Pfleging, Dagmar Kern, and Albrecht Schmidt. 2011. 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)*. Association for Computing Machinery, New York, NY, USA, 71–78. https://doi.org/10.1145/2381416.2381428 event-place: Salzburg, Austria.

[84] Galit Shmueli. 2010. To Explain or to Predict? *Statist. Sci.* 25, 3 (Aug. 2010). https://doi.org/10.1214/10-STS330

[85] John Sweller. 1994. Cognitive load theory, learning difficulty, and instructional design. *Learning and Instruction* 4, 4 (1994), 295–312. https://doi.org/10.1016/0959-4752(94)90003-5

[86] Trent Victor, Marco Dozza, Jonas Bärgman, Christian-Nils Boda, Johan Engström, and Gustav Markkula. 2015. *Analysis of naturalistic driving study data: Safer glances, driver inattention, and crash risk*. Technical Report SHRP 2 Report S2-S08A-RW-1.

[87] Frederik Diits Vikström. 2022. Physical buttons outperform touchscreens in new cars, test finds | Vi Bilägare. https://www.vibilagare.se/english/physical-buttons-outperform-touchscreens-new-cars-test-finds#vote

[88] Michael L. Watkins, Ivan A. Amaya, Paul E. Keller, Michael A. Hughes, and Edward D. Beck. 2011. Autonomous detection of distracted driving by cell phone. In *2011 14th International IEEE Conference on Intelligent Transportation Systems (ITSC)*. IEEE, Washington, DC, USA, 1960–1965. https://doi.org/10.1109/ITSC.2011.6083026

[89] Emmet White. 2023. Infotainment Screens Actually Distract Drivers, Study Shows. https://www.autoweek.com/news/technology/a42384978/infotainment-screens-make-distracted-drivers-swedish-study-finds/

[90] Chris Wickens, Jason McCarley, Lisa Thomas, MID Foyle, A Goodman, and BL Hooey. 2003. Attention-situation awareness (A-SA) model. In *NASA aviation safety program conference on human performance modeling of approach and landing with augmented displays*. 189.

[91] John R. Wilson and Sarah Sharples (Eds.). 2015. *Evaluation of Human Work* (0 ed.). CRC Press. https://doi.org/10.1201/b18362

[92] Claes Wohlin. 2014. Guidelines for snowballing in systematic literature studies and a replication in software engineering. In *Proceedings of the 18th International Conference on Evaluation and Assessment in Software Engineering*. ACM, London England United Kingdom, 1–10. https://doi.org/10.1145/2601248.2601268

[93] Bertram Wortelen. 2013. Simulating Attention Distribution of a Driver Model: How to Relate Expectancy and Task Value? (2013).

[94] Rachael A. Wynne, Vanessa Beanland, and Paul M. Salmon. 2019. Systematic review of driving simulator validation studies. *Safety Science* 117 (2019), 138–151. https://doi.org/10.1016/j.ssci.2019.04.004

[95] Lirong Yan, Yi Wang, Changhao Ding, Mutian Liu, Fuwu Yan, and Konghui Guo. 2019. Correlation Among Behavior, Personality, and Electroencephalography Revealed by a Simulated Driving Experiment. *Frontiers in Psychology* 10 (July 2019), 1524. https://doi.org/10.3389/fpsyg.2019.01524

[96] Kristie Young and Michael Regan. 2007. Driver distraction: A review of the literature. *Distracted driving* 2007 (2007), 379–405. https://citeseerx.ist.psu.edu/document?repid=rep1&type=pdf&doi=5673cc6c48ed46e3a2c83529e0961c83a3710b9a Publisher: Australasian College of Road Safety.

[97] Richard Young. 2017. Predicting Relative Crash Risk from the Attentional Effects of the Cognitive Demand of Visual-Manual Secondary Tasks. 2017–01–1384. https://doi.org/10.4271/2017-01-1384

[98] Sunbola Zatmeh-Kanj and Tomer Toledo. 2021. Car Following and Microscopic Traffic Simulation Under Distracted Driving. *Transportation Research Record: Journal of the Transportation Research Board* 2675, 8 (Aug. 2021), 643–656. https://doi.org/10.1177/03611981211000357

[99] Xiang Zhang, Hans-Frederick Brown, and Anil Shankar. 2016. Data-driven Personas: Constructing Archetypal Users with Clickstreams and User Telemetry. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*. ACM, San Jose California USA, 5350–5359. https://doi.org/10.1145/2858036.2858523

[100] Wenshuai Zhao, Jorge Pena Queralta, and Tomi Westerlund. 2020. Sim-to-Real Transfer in Deep Reinforcement Learning for Robotics: a Survey. In *2020 IEEE Symposium Series on Computational Intelligence (SSCI)*. IEEE, Canberra, ACT, Australia, 737–744. https://doi.org/10.1109/SSCI47803.2020.9308468