Prototyping Autonomous Vehicle Windshields with AR and Real-Time Object Detection Visualization: An On-Road Wizard-of-Oz Study

Lukas A. Flohr
lukas.flohr@ergosign.de
Ergosign GmbH
Munich, Germany
Saarbrücken Graduate School of Computer Science
Saarland Informatics Campus
Saarbrücken, Germany

Joseph S. Valiyaveettil
josephsv96@gmail.com
Ergosign GmbH
Munich, Germany

Antonio Krüger
krueger@dfki.de
German Research Center for Artificial Intelligence (DFKI)
Saarland Informatics Campus
Saarbrücken, Germany

Dieter P. Wallach
dieter.wallach@hs-kl.de
University of Applied Sciences Kaiserslautern
Kaiserslautern, Germany
Ergosign GmbH
Saarbrücken, Germany

Figure 1: On-road wizard-of-oz prototyping in an electric minivan (i) with a TV mounted on the headrests of the front seats (iii) that displays a video stream from the vehicle's windshield with AR-based real-time object detection visualization (ii) provided by an embedded computing platform (Nvidia Jetson Nano).

ABSTRACT

Autonomous vehicles (AVs; SAE levels 4 and 5) face substantial challenges regarding acceptance and UX. Novel human-machine interfaces (HMIs) providing transparent system information could account for those and facilitate adoption. However, since the availability of AVs for early concept studies is limited, context-based interface prototyping is required. This paper demonstrates the prototype and wizard-of-oz-based on-road evaluation of a futuristic windshield HMI concept that visualizes real-time object detections via augmented reality (AR). In a mixed-methods within-subjects study ($N = 30$), participants assessed three early-stage concept variants to explore whether object detection visualization can counteract the aforementioned challenges. The findings confirm that transparent system feedback can increase understandability, perceived usefulness, and hedonic UX, but the amount and the timing of the provided information are crucial. The applied prototyping method proved suitable for investigating HMI concepts with real-time AR on urban roads. Based on a critical discussion, the paper concludes with design and prototyping recommendations.

CCS CONCEPTS

• **Human-centered computing** → **HCI design and evaluation methods**; **Empirical studies in HCI**; *User studies*; *Interface design prototyping*; • **Computing methodologies** → *Mixed / augmented reality*; *Computer vision.*

KEYWORDS

autonomous vehicles, augmented reality, wizard-of-oz, acceptance, user experience, object detection, human-centered artificial intelligence, computer vision, computational interaction, visualization, windshield interface, context-based prototyping.

DIS '23, July 10–14, 2023, Pittsburgh, PA, USA

ACM ISBN 978-1-4503-9893-0/23/07.
https://doi.org/10.1145/3563657.3596051

ACM Reference Format:
Lukas A. Flohr, Joseph S. Valiyaveettil, Antonio Krüger, and Dieter P. Wallach. 2023. Prototyping Autonomous Vehicle Windshields with AR and Real-Time Object Detection Visualization: An On-Road Wizard-of-Oz Study. In *Designing Interactive Systems Conference (DIS '23), July 10–14, 2023, Pittsburgh, PA, USA.* ACM, New York, NY, USA, 15 pages. https://doi.org/10.1145/3563657.3596051

1 INTRODUCTION

Machines are taking over more and more tasks that humans previously performed – this process is called automation [54]. To describe the degree of automation in road vehicles, the Society of Automotive Engineers (SAE) [58] classifies six levels from 0 to 5. Throughout these levels, vehicle automation aims to increase safety, efficiency, and comfort, e.g., in the form of driver assistance systems [2]. Automated vehicles are expected to reduce traffic jams [64] and lower air pollution [67]. While in SAE levels 0 – 3, a human driver is required to perform the primary driving tasks (i.e., steering and acceleration) or at least for parts of it, this is not the case for levels 4 and 5 [58]. This means that vehicles with levels 4 and 5 are required to handle all traffic situations that might occur in their operational design domain [58]. Considering users' mental models [75], we use the term autonomous vehicles (AVs) to refer to SAE levels 4 and 5.

In AVs, the role of humans entirely shifts to passive passengers without control over the primary driving tasks. Consequently, passengers need to accept this unfamiliar and potentially awkward situation of being exposed to an artificial intelligence (AI) powered system's actions and decisions. Related work identified trust as a critical challenge for acceptance [36] as well as further "concerns about safety, security, usability, accessibility, and comfort" [55]. For successful adoption of the technology, those challenges must be addressed [36]. Since vehicle automation is becoming more complex and interconnected, Lacher et al. [45] conclude that a clear understanding of people, systems, and their interaction in a particular environment is required. Human-centered AI (HCAI) [56, 63] can provide an adequate framework and mindset to achieve this goal. Besides understanding humans, their abilities, and needs, an essential aspect of HCAI is "to help humans understand AI systems" [56]. Transparent communication – which, in this context, is about providing users access to the data and workflows inside an AI-based system [56] – may provide the basis for such an understanding [56]. Eventually, it may increase people's confidence and willingness to use these systems [56]. Related work indicates that contextual details and information on the AI-powered systems' status, reasoning, and actions could affect AV passengers' perceived safety, trust, acceptance, and UX [13, 22, 42, 47, 48, 53, 72]. Based on available AV sensor data (e.g., detected objects), human-machine interfaces (HMIs) could provide such information to supply passengers with transparent and understandable explanations of system behavior. These might be able to compensate (to some part) for the absence of a human driver and counteract said challenges.

The paper's contribution to this field is twofold. *First*, we investigate whether AV passengers' acceptance and UX can be increased by providing transparent information on the AI-powered system's reasoning with the (computer-vision-based) visualization of detected objects and how this information should be displayed during the ride. We conducted an on-road wizard-of-oz (WoOz) study ($N = 30$) in a real, urban environment and compared three early-stage concept variants for a futuristic windshield interface: (1) a baseline concept without object detection against two forms of real-time visualizations: (2) an unobtrusive status bar with counts of detected objects per class, and (3) salient AR overlays. The findings from the mixed-method real-world driving study contribute to prototyping and designing suitable passenger information systems for AVs that can counteract acceptance hurdles and support positive UX. *Second*, we demonstrate a straightforward WoOz-based prototyping approach to investigate real-time information visualization with a futuristic AR windshield prototype, which also served as an enabler for the described study. For this purpose, we combined an embedded AI system with object detection algorithms with an easily reproducible WoOz setup. We documented the approach in this paper to serve future work as a foundation for context-based interface prototyping and evaluation of AV HMI concepts in real-world scenarios and as inspiration for researchers and practitioners.

2 RELATED WORK

Situated within the broader context of vehicle automation and human-centered AI, this paper is primarily concerned with (1) the acceptance and UX challenges of AVs, (2) the development of adequate in-vehicle HMIs capable of counteracting these challenges, and their exploration and evaluation using (3) context-based prototyping.

2.1 Acceptance and UX Challenges

Apart from achieving technological maturity, AVs face significant challenges in terms of future users' acceptance [36, 51, 55]. Since people need to give up control and choice to an autonomous system, Kaur and Rampersad [36] identified public trust as the primary adoption barrier of AVs and conclude that the vehicles' reliability and the match with users' performance expectations are crucial adoption factors. Privacy concerns (e.g., in terms of surveillance and tracking), as well as security concerns (e.g., in terms of software errors or hacker attacks), were mentioned as further trust determinants [36]. When AVs become integrated into public transportation (PT) systems, rides will be shared with others: this raises further security concerns among potential users related to the interactions with strangers without human oversight (e.g., through a bus driver) [62]. Additionally, potential users carry concerns about safety, usability, accessibility, and comfort [55]. Based on a comprehensive understanding of users, system, and their environment [45], human-computer interaction needs to counteract these acceptance and UX challenges to support the technology's adoption. Toward this understanding, Chen [11] proposed an extension of the technology acceptance model (TAM; [18]) for the AV domain and concluded that peoples' attitudes and perceived enjoyment directly affect people's intention to use AVs. At the same time, trust, perceived ease of use, and perceived usefulness affect people's attitude toward the technology [11].

2.2 In-Vehicle Human-Machine Interfaces

Since no human driver is required in SAE levels 4 and 5, HMIs remain the sole touchpoint of passengers and the AI-powered system. They, therefore, take a crucial role in counteracting acceptance challenges and fostering adequate trust. HMIs for interacting with AVs can manifest as mobile booking and companion applications, terminals at mobility hubs, or HMIs inside and outside the vehicle. Jansen et al. [33] provided a comprehensive overview of the in-vehicle design space for input and output modalities and information locations. Their systematic literature review revealed that the most established modalities for human-vehicle interaction are visual, auditory, kinesthetic, and tactile [33].

To support passengers in understanding AI-powered AVs as well as their intentions and actions, transparent internal communication via the vehicle's infotainment and automation HMIs [4] could be the key for high acceptance and positive UX. In this context, related work on interacting with automated vehicles with lower automation levels (i.e., up to SAE level 3) suggested providing users with information on the current system status and the driving context in order to explain the automated system's decisions and actions to its occupants [14, 22, 42, 47]. Supplying users with information on surrounding elements can increase users' situation awareness [13, 47] and trust [14, 15, 53, 72, 73].

Oliveira et al. [53] found in an indoor study with an experimental level 4 vehicle that providing transparent system information via HMIs can increase trust. In their comparison of HMI concepts, an AR-based variant received the best assessment [53]. Similar results have also been reported for AR interfaces in vehicles with lower levels of driving automation, e.g., [47, 72, 73]. Colley et al. [13] investigated the potential of semantic segmentation visualization of detected objects to increase drivers' trust and situation awareness in vehicles with conditional driving automation (SAE level 3) with two online studies. Their results indicate the potential of (AR-based) visualization to increase situation awareness but do not reveal significant effects on driver's trust [13]. In a consecutive work, Colley et al. [15] found (AR-based) visualizations related to situation prediction are perceived negatively and degrade the attributed capabilities of the automated vehicle. However, they stated that transparent information visualizations could serve as a measure to calibrate trust. In line with that, Wintersberger et al. [72, 73] concluded, based on a simulator study, that traffic augmentations can increase drivers' trust in ambiguous situations (e.g., dense fog) and in automated driving systems in general. However, not all passengers might want to have such information at all times [53]. Thus, the design and amount of provided information and explanation are crucial since "more information does not necessarily lead to more trust" [48]. Similar results are also observable in other domains: e.g., Kizilcec et al. [40] found that making an algorithmic interface for peer assessment more transparent by providing explanations can increase trust but also diminish already built confidence if too much information is provided.

2.3 Context-Based Prototyping with Wizard-of-Oz

While driverless rides become more and more experience-able through the increasing deployment of AVs on test tracks [21, 51] and also on public domains [29, 31, 52, 57], driverless rides in complex urban environments remain limited and mostly only feasible under restricted conditions (e.g., in terms of specific test tracks and scenarios, speed limitations, legal regulations, or requiring constant human supervision). Consequently, the design and conduct of empirical concept studies with such are limited. Context-based prototyping [23, 25, 32] can provide suitable approaches to overcome this hurdle and help to consider both the complex context and the experience of an autonomous ride from the early development phases. Despite real-world AVs, popular methods in the automotive domain are simulators and WoOz setups. Depending on a study's focus, experimenters must weigh the pros and cons of the respective methods and setups [25]. Simulators immerse study participants in a virtual environment by using either computer-generated imagery (CGI) [22, 28, 73] or (immersive) video [23, 24, 27, 43, 44]. While simulators enable the consideration of various scenarios with high controllability and reproducibility [19, 59], it is still challenging to create high-fidelity representations of complex environments (e.g., urban driving scenarios). Furthermore, simulators are restricted to an artificial lab context.

In contrast to simulators, WoOz can be applied to prototype AVs and their HMIs in real-world environments, i.e., on public roads [20, 35, 39, 49, 69]. Over the past decade, the method's popularity has enormously increased within the automotive domain, e.g., to evaluate new HMI concepts [30] or to investigate non-driving related activities [20], leading to the proclamation of the "renaissance" of WoOz [3]. It enables automated system evaluation prior to their actual availability [3] and can go beyond the limitations of laboratory contexts [69]. The basic idea of WoOz is to make participants believe that they are interacting with an intelligent and/or automated system while humans do in fact simulate it – the so-called *wizards* [6]. When using the method to prototype AVs, study participants need to believe that the vehicle is driving automated while a hidden human driver – the driving wizard – controls it [3]. Bengler et al. [3] provided an overview of typical WoOz vehicle setups, which depend on the automation level of the tested system and the degree of the participant's (illusion of) control and form of input. Given that AV passengers do not need to control the vehicle, (mock-ups of) steering wheels and pedals are not required for study participants. Typical setups place participants either on the co-driver's seat [1, 38, 39, 69] or in the back [20, 35, 49] and separate them from the wizards. For instance, Karjanto et al. [35] and Detjen et al. [20] placed participants in the back and separate them with a wall and a mounted TV that displays a video stream of a camera at the vehicle's windshield.

Besides the advantages regarding low contextual limitations, WoOz poses methodological challenges. With respect to a study's validity, it is essential to make participants believe in the WoOz illusion and to have the simulated automation behave like an actual automated system would [50, 69]. For the latter, driving wizards need to be instructed on the desired driving style (e.g., smooth and conservative, like "a professional limo driver" [1]). To ensure reliability of the study, this driving style must be consistently reproduced by the driving wizard throughout all sessions [50]. If multiple driving wizards are used, each needs to be able to recreate the same style to ensure objectivity [50]. To initiate and keep up the illusion, *cover stories* [20] are used to tell participants about the (simulated) capabilities of the automated system – e.g., driving autonomously on urban streets. Apart from the methodological challenges, variations in environmental factors such as traffic density, presence and behavior of other road users, weather, and lighting conditions pose further hurdles and might impact the WoOz study goal, its

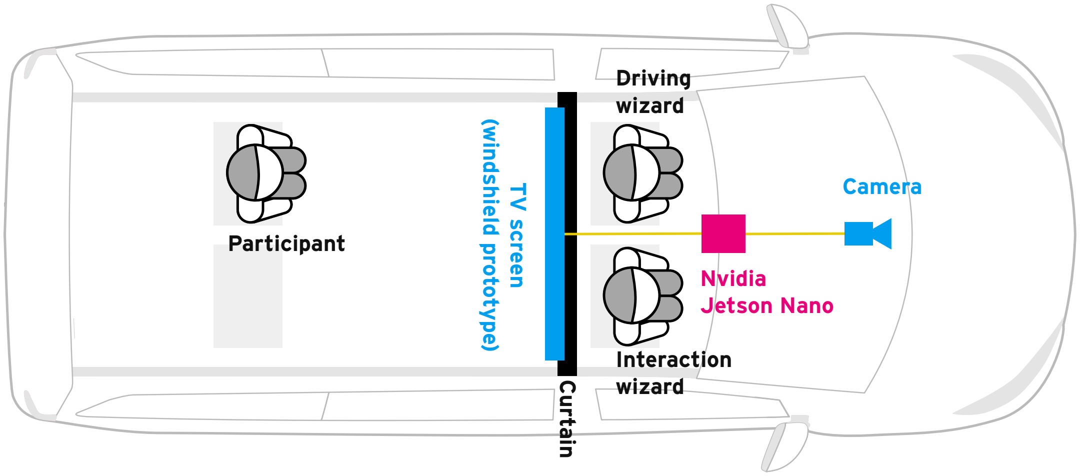

Figure 2: Schematic illustration of the used wizard-of-oz-based prototyping setup.

reliability [50], and the comparability of test rides [3]. Therefore, [3] proposed to test not only hypotheses in terms of the research questions but also the "comparability of test drives and the believability of the illusion".

3 ON-ROAD WIZARD-OF-OZ STUDY

Related work reveals the potential of providing explanations and transparent system feedback to increase acceptance, trust, and UX of automated vehicles [13, 15, 22, 42, 47, 48, 53, 72]. Most of these previous studies (except [53]) investigate vehicles with lower levels of driving automation where a human driver is still required (i.e., up to SAE level 3) and focus on providing system feedback for specific situations (e.g., maneuvers in ambiguous situations [73]). Furthermore, they were conducted online [13, 15], in labs with simulated artificial environments [22, 42, 47, 48, 72], or on restricted (in-door) test tracks [53], but not on real urban roads. The questions arise about whether (1) previous findings from lower automation levels can be transferred to driverless AVs and (2) to dynamic and complex urban real-life environments, and (3) whether, when, and how transparent information and explanations should be displayed in AVs. We address the identified research gap by investigating the following research questions (RQs) in an empirical user study.

- **RQ1** Can we increase AV passengers' acceptance and UX by providing transparent system information via (AR-based) visualization of detected objects in the vehicle windshield?
- **RQ2** How and when should this information be displayed during AV rides in urban environments?
- **RQ3** How can we create a suitable prototyping framework to investigate RQ1 and RQ2, as well as related questions in complex urban real-life environments?

3.1 Study Design and Prototyping Framework

We adopted a within-subjects design to achieve high internal validity and to minimize the effects of random noise [10], e.g., caused by varying environmental factors. To investigate the effects of real-time object detection visualization on passengers in a natural urban environment (i.e., in real traffic), we created a contextualized prototyping framework based on a WoOz setup in combination with a prototype for a futuristic windshield HMI implemented on an embedded computing board (Nvidia Jetson Nano). Before conducting the study, its design, setup, procedure, and data collection were assessed by the Ethical Review Board of Saarland University with the process number 21-11-4. The board did not raise any ethical concerns. Furthermore, the study was conducted in accordance with applicable ethical principles stated in the Declaration of Helsinki [74].

3.1.1 Wizard-of-Oz Setup and Prototyping Considerations. Due to the reasons elaborated in Section 2.3, we opted for a WoOz approach as a basis for the study and the investigation of our RQs. Inspired by previous works, especially by Karjanto et al. [35] and Detjen et al. [20], we created a straightforward WoOz setup (Fig. 1 and 2) that we used as an on-road simulation of an AV ride through the city. An electric minivan (Mercedes-Benz EQV) served as a basis for the setup. The car came with a modern appearance and offered sufficient space for the setup. Since we used a rental car, we aimed – in contrast to previous works – to create an easily deconstructable setup without the need for physical adjustments (e.g., drill holes) that can also be easily reproduced in similar vehicles. To achieve this, we mounted a TV (Hisense 43" 4K) at the front seats' headrests using a wooden board with a standard TV wall mount and screw pipe clamps. To provide the basis for the investigation of RQ1 and RQ2 and as a potential answer to RQ3, we connected the TV to an embedded computing platform (Nvidia Jetson Nano). The Jetson displayed the HMI prototype (Section 3.1.2), including the video stream of a consumer webcam (Logitech BRIO) mounted in the vehicle's windshield. We then mounted a black curtain with heavy-load magnets and duct tape on the car's ceiling to separate the vehicle's front and back parts. A power inverter (NDDI 600 W) inverted the vehicle's 12 V DC power plug to 240 V AC to power the TV and the Jetson. For safety measures, we added an additional socket with surge protection. Based on the consultation of an automotive expert witness auditing company, we made some

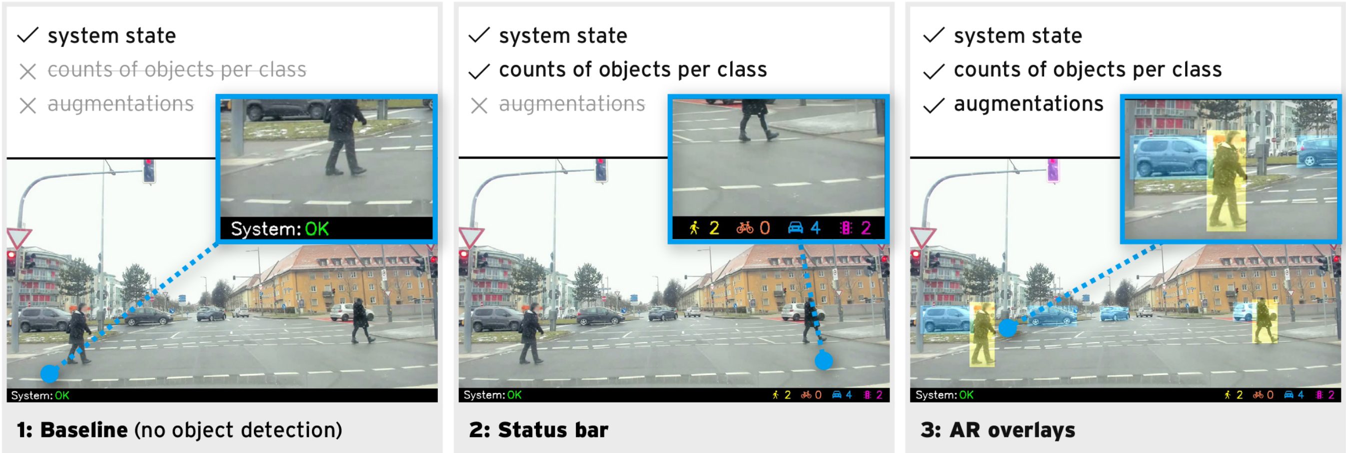

Figure 3: Overview of the three concept variants displayed in the windshield prototype.

final adjustments and optimizations by better securing the load and setup. Lastly, we added a foiling to the vehicle's exterior that marked the vehicle as a research vehicle to support the WoOz cover story used.

Following the recommendations of related work, we instructed the wizard to perform a conservative and relaxed driving style, like "a professional limo driver" [1]. To increase objectivity, all sessions were driven by the same experienced driver who was familiar with the vehicle (familiarization time of 3 weeks prior to the study) and aimed to reproduce the same driving style throughout all sessions consistently. As the WoOz setup limited the view out of the vehicle, the co-driver (interaction wizard) supported the driver in difficult situations during the test rides, e.g., by spotting vulnerable road users when turning right.

3.1.2 Windshield HMI Prototype. The futuristic (AR-capable) windshield HMI prototype was implemented as a graphical user interface (GUI) application displaying image frames from a webcam using OpenCV. Depending on the concept variant – the prototype draws real-time AR bounding box visualizations over detected objects and/or shows a descriptive status bar with counts of objects per class for the detections (Fig. 3). The detector uses a pre-trained model (YOLOv4 [7] trained on the COCO dataset [46]) optimized for inference using ONNX and TensorRT and runs on the Jetson's GPU. Reducing the complexity of the HMI and the study, the application merges object classes from the dataset into four main headers: pedestrians, cyclists, vehicles, and traffic signs. To reduce latency and jitter from object visualizations and increase the frame rate of the video feed, we implemented a periodical switch to a lower overhead object tracker that was periodically re-initialized by the object detector. The application was implemented on an Nvidia Jetson Nano embedded-computing board with a 4-core CPU, 4 GB RAM, and a 128-core GPU and displayed on the TV. We applied several optimization measures to display the video feed and the object visualizations with a fluent frame rate and sufficient resolution (TensorRT optimizations, joint detection and tracking). This resulted in a feasible resolution of 1280 x 720 pixels at about 24 fps, that was, with regards to the passengers' viewing distance of about 160 cm, sufficient. We want to note that the early computer-vision-based prototype's performance has limits and is not up to the accuracy and precision of cutting-edge sensing systems, e.g., [57, 70]. Nevertheless, the implementation provides a suitable and flexible prototyping basis to investigate our research questions at an early development stage.

For the design of the AR-based object visualizations, we adopted two-dimensional bounding box overlays as they are widely used in the computer vision domain for basic object annotation (Fig. 3: 3). Depending on the object class (e.g., pedestrian) , the overlays had different colors (e.g., yellow). In the design phase, we also considered approaches and visualization techniques, such as 3-dimensional AR markers or as representations on a separate display, as well as combinations with "classic" information, feedback, and navigation concepts (e.g., displaying the planned route on a map). However, the reported study was intended as an early concept study, which is why we focused on (AR-based) object visualizations. To investigate their general potential, we created a baseline variant of the prototype without feedback on detected objects (Fig. 3: 1). Since related work pointed out that the amount of displayed information might affect passengers' experience [53], we created an intermediate variant which visualized detected objects as counts per object class in a status bar only (Fig. 3: 2). The variants are designed sequentially. I.e., variant 3 also includes the status bar of variant 2. Furthermore, all three variants displayed general information on the overall system state ("System: OK"), which provided passengers with baseline information on the system's functionality throughout the variants. We opted to provide this baseline information for two reasons: to inform (and convince) passengers that the simulated AV is driving autonomously and to ensure them that everything is fine – even if there is no further information displayed. In the conducted study, the system state never changed.

3.2 Participants

With a sample of 30 participants (14 female, 16 male, 0 diverse, 0 n/a) between the ages of 20 and 70 ($M = 37.6, SD = 11.9$), we achieved a statistical power of .84 (calculated with G*Power 3.1) for the calculation of inferential statistics (repeated measures analysis

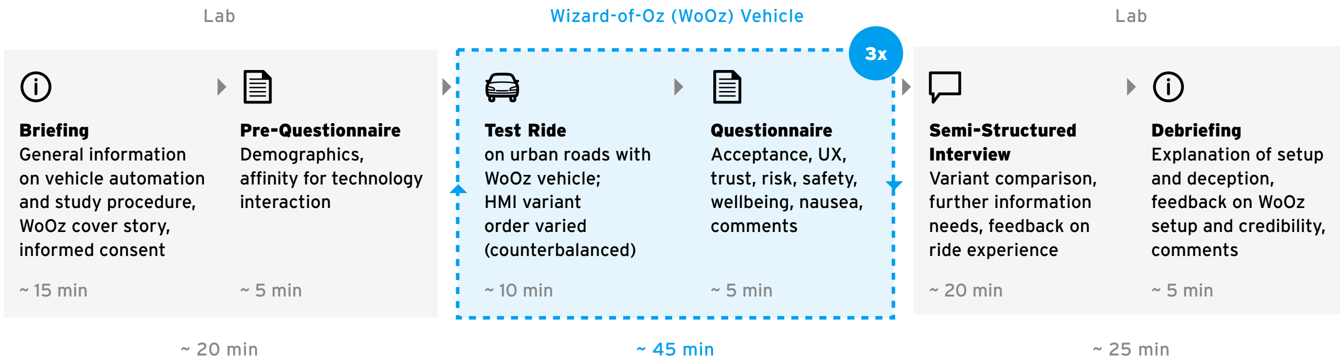

Figure 4: Study procedure of the on-road wizard-of-oz study.

of variance (RM-ANOVA) with within factors and three measurements) assuming medium effects according to Cohen [12] and an alpha error rate of $\alpha \leq .05$. The sample had a medium-high affinity for technology interaction (ATI short scale [71]: $M = 4.4, SD = 1.3; min = 0, max = 6$) and was well educated (highest degree: 19 with university degrees, six with advanced school-leaving certificates, three with intermediate school-leaving certificates, two with other degrees). Three participants had an uncorrected visual impairment (two myopia, one red-green color blindness), which they reported not having posed a problem during the study. All participants were external from our institution and recruited via online postings, mailing lists, and advertising posters. Each participant received financial compensation of 30 €.

3.3 Procedure and Data Collection

Each participant took part in an individual session with an experimenter and a note taker, who also took over the roles of the driving and interaction wizards during the test rides. The sessions took about 90 min and were structured into three main phases (Fig. 4): (1) briefing and pre-questionnaire in the lab, (2) test rides and consecutive questionnaire in the WoOz vehicle, and (3) semi-structured interview and debriefing in the lab. Following a mixed-method approach [17], we collected both qualitative and quantitative data. For an in-depth post hoc analysis, we recorded audio during the rides and the interviews and took notes during the sessions. For the quantitative assessment and comparison of the HMI variants, we used standardized UX, trust, and acceptance questionnaires and single-items to assess perceived risk, safety, wellbeing, and nausea during the rides (Table 1).

3.3.1 Briefing and Pre-Questionnaire. At the beginning of the study session, the participant received a detailed briefing on the study's purpose and procedure. This was already initialized our WoOz deception. As a part of our cover story, we explained the basics of autonomous driving technology and automation levels. We told participants that we would conduct the test rides with an actual AV capable of handling all driving situations but requiring the presence of a safety driver (the driving wizard) due to current legal regulations. Furthermore, we declared that we wanted to evaluate futuristic windshield HMI concepts that are technically not yet feasible to be implemented in the vehicle. This served as the explanation for the TV-based prototyping. By providing passengers with this information, we aimed to shift the focus toward the HMI prototype and away from the WoOz setup. Furthermore, we explained to participants how the AV's object detection works and that AVs use it to navigate safely through traffic. We outlined that some of the tested HMI concepts might provide this sensor information also to passengers to optimize their experience. It was added that the tested concepts are currently in an early prototyping phase and are, thus, using not the actual AV sensors but a single camera that we included in the AV for research purposes. Due to this early development stage, we explained that the system's performance is limited and might affect the correct display of the HMI information. After the briefing, participants signed an informative participation consent form and filled out a pre-questionnaire to provide information on their demographics and affinity for technology interaction (ATI-S [26, 71]).

3.3.2 Test Rides and Questionnaire. The test rides were conducted as a round-trip through an urban environment with two stops at parking lots and about 10 min driving time per variant. The variant order varied (counterbalanced) between sessions to decrease carry-over effects. Before starting the ride, participants were given some final notes on the setup. We encouraged them to think aloud and explained to them once again that they could pause or quit the study at any time without consequences. At the two stops, we changed the HMI variant and asked participants to fill out a digital questionnaire on a tablet to assess the respective HMI variant in terms of our dependent variables (Table 1). For the assessment of acceptance, we used the *Satisfying* and *Usefulness* scales of Van der Laan et al. (VdL) [68] and the scales *Perceived Enjoyment* and *Intention to Use* of Chen's TAM adaption [11]. As related work has identified trust as a key acceptance challenge for AVs [11], participants also assessed the variants in terms of trust and related factors using the scales *Trust in Automation*, *Reliability/Competency*, and *Understandability/Predictability* by Körber [41]. For the assessment of *pragmatic and hedonic UX*, we used the short version of the User Experience Questionnaire (UEQ-S) [61]. In addition, we used single-item scales to let participants assess perceived *risk* ("I considered the ride risky."), *safety* ("I felt safe during the ride."), *wellbeing* ("I felt comfortable during the ride."), and *nausea* ("I felt

Table 1: Dependent variables and their operationalization.

	Scales	Items	Reference
Acceptance	Satisfying	4 bipolar items; 5-point scale	[68]
	Usefulness	5 bipolar items; 5-point scale	[68]
	Perceived Enjoyment	3 items; 5-point Likert-type scale	[11]
	Intention to Use	2 items; 5-point Likert-type scale	[11]
Trust	Trust in Automation (TiA)	2 items; 5-point Likert-type scale	[41]
	Reliability/Competency (R/C)	6 items; 5-point Likert-type scale	[41]
	Understandability/Predictability (U/P)	4 items; 5-point Likert-type scale	[41]
UX	Pragmatic UX	4 bipolar items; 7-point scale	[61]
	Hedonic UX	4 bipolar items; 7-point scale	[61]
Risk & Safety	Risk	single-item; 5-point Likert-type scale	
	Safety	single-item; 5-point Likert-type scale	
Wellbeing	Wellbeing	single-item; 5-point Likert-type scale	
	Nausea	single-item; 5-point Likert-type scale	

nauseous during the ride."). Participants could comment on their assessments via free-text input fields. Following the recommendation of [3] to collect environmental data of the test rides, experimenters documented weather conditions and traffic density.

3.3.3 Semi-Structured Interview and Debriefing. After the test rides, we conducted a semi-structured interview using closed and open questions. We recapitulated the rides and HMI variants and talked to participants about what they liked and disliked, their preferences, and what they would suggest for future systems. We also asked participants which variant they liked best and why. At the end of the interview, we lifted the WoOz deception and explained the reasons. After the explanation, we asked participants the *WoOz control question* ("Did you believe that the vehicle was driving autonomously?") to directly assess the deception's effectiveness.

4 RESULTS

For the quantitative results, we used *JASP 0.16* [34] and *jamovi 2.2.5* [66] to calculate descriptive and inferential statistics. In a second step, we analyzed the qualitative data from the interviews, ride recordings, and questionnaires. All recordings were transcribed using the speech-to-text function of *Condens* [16], reviewed, and manually optimized afterward. Following an inductive thematic analysis approach [8, 9], three researchers worked collaboratively. We used Condens and a digital *Miro* whiteboard to analyze and structure the data in order to identify patterns that describe essential information concerning our research questions. Each researcher started with analyzing a few sessions and derived an initial set of codes which was then reviewed by the others and merged to create a joint codebook. The codebook and coding fragments were iteratively refined throughout the analysis. Finally, the thematic analysis was complemented with the questionnaire results and session notes.

4.1 Dependent Variables

Besides a descriptive analysis, we conducted repeated measures analysis of variances (RM-ANOVAs) to search for statistically relevant effects. For the interpretation of calculated effect sizes, we refer to Cohen [12]. If one or multiple assumptions of the RM-ANOVAs (independence, normality, sphericity) was found violated for a particular scale, we calculated non-parametric Friedman tests and Conover's post hoc comparisons.

4.1.1 Acceptance. Results of the VdL acceptance questionnaire [68] show medium ratings of *Satisfying* and *Usefulness* scales with the Baseline achieving highest ratings with regard to *Satisfying* and the AR overlays highest in terms of *Usefulness* (Fig. 5). While no significant difference was found for *Satisfying* ($F(2, 58) = 1.590, p = .213, \eta^2_G = 0.030$), a significant medium effect was found for *Usefulness*, $F(2, 58) = 7.881, p < .001, \eta^2_G = 0.136$). Post hoc tests revealed significantly better *Usefulness* ratings of the AR overlays compared to the Baseline ($t = 3.806, p_{holm} = .001$) with a medium-sized effect of $Cohen's\ d = 0.695$ and compared to variant 2 (status bar with counts; $t = 2.882, p_{holm} = .011$) with a medium-sized effect of $Cohen's\ d = 0.526$. Regarding the *Enjoyment* scale of Chen's TAM adaption, all variants achieved high ratings (Fig. 5) with no meaningful effect ($F(2, 58) = 0.925, p = .402, \eta^2_G = 0.014$). Similarly, all variants achieved medium-high ratings for *Intention to Use* (Fig. 5) without relevant differences, $F(2, 58) = 1.553$, $p = .225$, $\eta^2_G = 0.020$.

4.1.2 UX. With regard to *pragmatic UX quality*, all three variants received above middle ratings (Fig. 5) with no significant differences between them, $F(2, 58) = 1.590, p = .213, \eta^2_G = 0.030$. For *hedonic UX quality*, larger deviations ranging from above middle ratings (AR overlays) to medium-low ratings (baseline, status bar; Fig. 5) with a significant large effect were found, $F(2, 58) = 10.447$, $p = .001$, $\eta^2_G = 0.169$. Post hoc tests show significant higher hedonic quality with the AR overlays compared to the baseline ($t = 4.334, p_{holm} < .001$) with a medium-sized effect of $Cohen's\ d = 0.791$ and compared to variant 2 ($t = 3.136, p_{holm} = .005$) with a medium-sized effect of $Cohen's\ d = 0.572$.

4.1.3 Trust. The HMI variants received medium to medium-high assessments regarding *Understandability/Predictability* (Fig. 6) with significant differences between the variants showing a medium-sized effect, $F(2, 58) = 8.128$, $p < .001$, $\eta^2_G = 0.108$. Post hoc tests revealed significantly better *Understandability/Predictability*

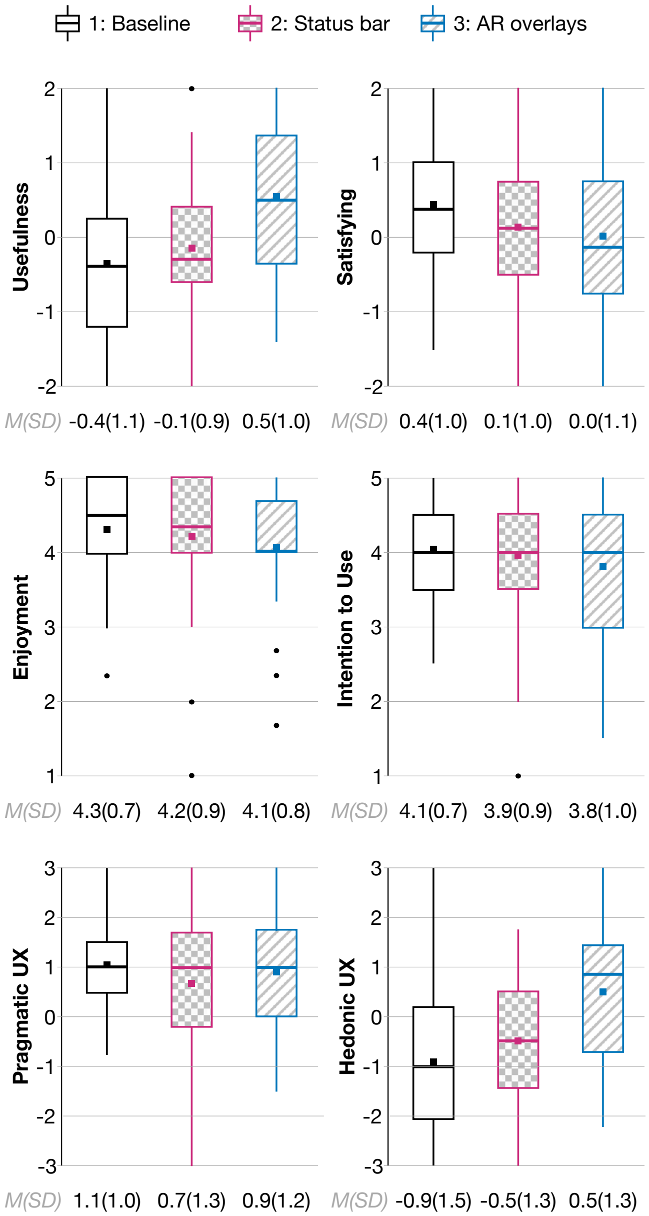

Figure 5: Boxplots and $M(SD)$ of acceptance and UX scales for the three concept variants.

of the variant with AR overlays compared to the Baseline variant ($t = 3.810$, $p_{\text{holm}} < .001$) with a medium-sized effect of $Cohen's\ d = 0.696$ and compared to variant 2 ($t = 3.048$, $p_{\text{holm}} = .007$) with a medium-sized effect of $Cohen's\ d = 0.556$. Regarding *Reliability/Competency*, the variants obtained above middle ratings (Fig. 6) without meaningful differences, $F(2, 58) = 2.309$, $p = .108$, $\eta^2{}_G = 0.025$. Similarly, all three variants received medium, above middle ratings for overall *Trust in Automation* without a significant effect, $F(2, 58) = 1.803$, $p = .174$, $\eta^2{}_G = 0.019$.

4.1.4 Risk and Safety. *Risk* was rated low throughout all variants (Fig. 6) without meaningful differences, $\chi^2 = 1.869$, $p = .393$, $n = 30$. In accordance with that, the *Safety* scale received medium-high ratings in all conditions (Fig. 6) without significant differences, $F(2, 58) = 0.677$, $p = .512$, $\eta^2{}_G = 0.009$. The low risk values and the feeling of safety was often related to trust in the general capabilities of the automated system (e.g., P10: *"I am convinced of the capabilities of the system"*, P21: *"I trust the system. Unforeseen events were handled without problems."*) as well as the driving style (e.g., P16: *"[it] drives like me – safe"*; P24: *"The vehicle reacted with restraint in unusual situations. That was good."*; P28: *"Very relaxed way of driving [... and] good response of the vehicle to all situations."*).

4.1.5 Wellbeing and Nausea. With regard to *Wellbeing*, the single-item scale revealed positive assessments with medium-high ratings for the three variants (Fig. 6). While no significant difference between the variants was found ($\chi^2 = 3.774$, $p = .152$, $n = 30$), descriptive statistics suggest that participants felt slightly better using variants 1 or 2 (Fig. 6). This is similarly indicated by the *Nausea* scale (Fig. 6). While only a few participants reported Nausea symptoms, there is a significant difference between the variants, $\chi^2 = 7.357$, $p = .025$, $n = 30$. Nausea symptoms occurred significantly more often with AR overlays compared to variant 2, *Conover T-Stat* $= 2.838$, $p_{\text{holm}} = .019$. However, the differences between AR overlays and baseline (*Conover T-Stat* $= 1.845$, $p_{\text{holm}} = .140$) and between baseline and variant 2 (*Conover T-Stat* $= 0.993$, $p_{\text{holm}} = .325$) are not statistically significant. Four participants (P19, P23, P25, P30) also described motion sickness symptoms verbally. While P23, P25, and P30 related the symptoms to generally watching at the digital screen during the ride, P19 accounted them particularly to the AR overlays: *"I think if I drove here longer, I might feel a little dizzy [...] from the color fields."*. P30 added that wearing an FFP2 face mask during the ride further influenced the occurrence of the symptoms.

4.2 Qualitative Variant Assessment

Overall, variant 3 with the AR overlays was preferred by half of the sample. However, ten participants put the AR overlays on the last rank and ten rated variant 1 (baseline without information on detected objects) as their favorite. Only five participants preferred variant 2 with the object counts in the status bar. The following sections provide a detailed overview of the received qualitative feedback per variant.

4.2.1 Baseline (Variant 1). Seven of the ten participants that opted for variant 1 found the visualization of surrounding objects generally unnecessary (e.g., P12: *"because if you don't drive yourself anyway, then it doesn't really need to display anything."*). P10 considered only the general system feedback ("System: OK") relevant and the object visualizations as a *"gimmick [...] unless it really has the consequence to intervene"*. Two participants argued that less information is better when it comes to trust in the technology (e.g., P28: *"the system seems more trustworthy even though there is less information available"*).

4.2.2 Status Bar (Variant 2). Twenty participants considered the status bar unnecessary (e.g., P23: *"I found this nice, but somehow just not helpful."*) – in contrast to five participants who described the count display as helpful. While some mentioned that the object counts increased perceived safety ($n = 3$) and trust ($n = 3$;

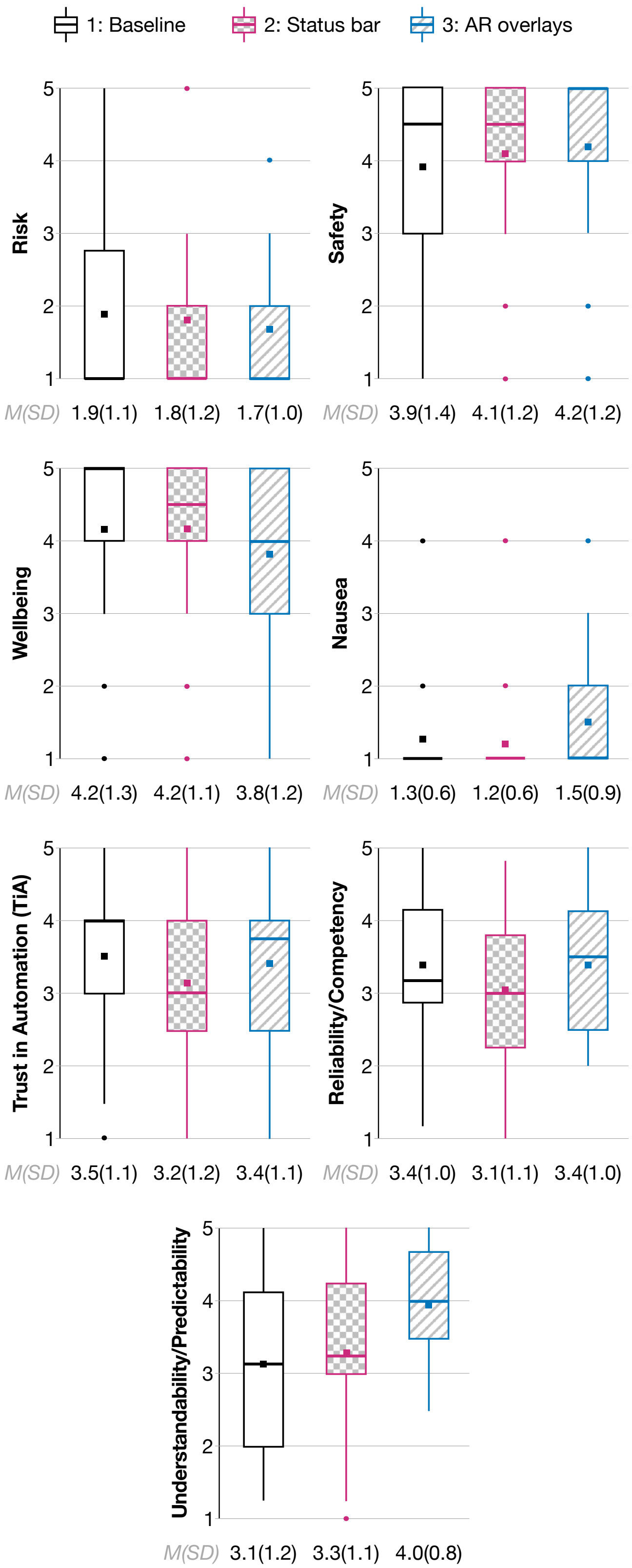

Figure 6: Boxplots and $M(SD)$ of trust, risk, safety, wellbeing, and nausea scales for the three concept variants.

e.g., P28: *"it's reassuring"*), others said it decreased both perceived safety ($n = 5$) and trust ($n = 3$). Since the counts jumped fast in some situations, eight participants perceived the fast refresh rate as irritating (e.g., P17: *"the display has made one restless [...] one is tempted to control the display"*). Without the matching overlays of variant 3, the meaning of the count display was unclear to 11 participants and left some with open questions and the desire for better contextualization of the information, e.g., P13: *"I would like to know how the car puts what it recognizes into context with the driving context"*.

4.2.3 AR Overlays (Variant 3). The AR overlays of variant 3 were considered to be helpful for (better) understanding the driving situation ($n = 11$; e.g., P23: *"You could see at a glance what was happening and classify it much better"*) and to build trust in the system ($n = 12$; e.g., P13: *"One could better understand the complexity of the system. Therefore, more trust"*). However, in 16 of the 30 sessions, participants described the AR overlays to be either annoying, irritating, or distracting (e.g., P17: *"You can't enjoy the ride"*; P21: *"Too many colored boxes"*; P26: *"somewhat annoying display",*). In some sessions, participants reported that the AR overlays decreased perceived safety ($n = 6$) and their trust in the system ($n = 3$). This ambivalence was further observable in the interviews, where many participants weighed the variants' pros and cons.

4.3 Visualization Design

Twelve participants desired to have only objects relevant to the current driving scene visualized. Regarding the visual design of the overlays and object counts, participants considered distinct colors for object classes to be useful ($n = 24$), as well as various visualizations for critical objects ($n = 20$; e.g., P18: *"if a pedestrian would run in, [the overlay] becomes red for example"*). On the other hand, eight participants would have generally preferred fewer colors.

4.3.1 Amount and Type of Object Visualization. Asking participants which objects they would want to get visualized, only four voted for all objects. Eleven participants wanted only objects marked that have an impact on the vehicle's ride, and 12 preferred only marking hazardous objects (e.g., P7: *"[would be] clearer"*; P26: *"All objects is too much and too confusing"*). While two participants remained undecided, only P19 opted for no object detection visualization at all (*"too distracting"*). The latter is particularly interesting considering that in the general assessment, nine more participants ranked variant 1 with no object visualization as their favorite. With regard to the question which kind of objects should be displayed, participants mentioned vulnerable road users to be most important (bicyclists: $n = 23$, pedestrians: $n = 24$). Vulnerable road users are considered more important than general obstacles ($n = 20$), own driving trajectory ($n = 20$), other vehicles ($n = 19$), traffic lights ($n = 17$), infrastructure ($n = 12$), traffic signs ($n = 10$), or street markings ($n = 6$). Most participants ($n = 16$) preferred visualizing detected objects according to their hazard level. P13 suggested overlays with a transparency level according to their relevance or criticality. In contrast, ten participants considered a special visualization for hazards unnecessary. Two participants pointed out that objects that are not visualized can be regarded as unrecognized and, therefore, be a safety risk. P14 and P24 asked whether the system could detect animals (e.g., dogs).

4.3.2 Configurability. In general, the sample majority ($n = 23$) argued for configurable display settings allowing passengers to choose what, when, and how information is displayed. E.g., P13: *"I think when you probably use the system more often [... the display] could be a distraction and I might want to turn it off"*; P30: *"I think it's good if everyone could decide for themselves"*). P5 suggested that the visualization should turn itself off automatically after a specific time but can be turned on again by the passengers. Five participants proposed that the UI screens should be usable for other things, e.g., as a second screen for mobile work, information, or entertainment.

4.3.3 Additional Ideas. P1 would have liked to use the augmentations at night to enable a kind of night vision for passengers. P15 wanted an onboarding tutorial explaining the displayed information and functionalities to first-time users or on-demand. Similarly, P30 would have wanted a more detailed legend that explains, e.g., the meaning of the colors. P14 would have found it helpful if an indicator for the object's moving direction gets displayed. P27 suggested acoustic warnings for critical situations so that passengers could prepare themselves, e.g., for occurring driving actions like emergency brakes. Many participants wished for improved (AR) visualization ($n = 6$) and suggested, e.g., not to display large overlays with "sharp" borders but rather, e.g., a spot or a dot (P28), a decent border (P4), a soft filling (P2), or a gradient or blur (P25), which they considered to be more convenient to look at and assumed to reduce flickering of the detection and consequent distractions.

4.4 Further Information Needs

A large part of the sample ($n = 14$) wanted to have information on the current location and the planned route, e.g., displayed on a map. Two participants suggested to have this on an extra display. Twelve participants would have liked to get location-based information about their surroundings, such as descriptions of landmarks. Seven participants wanted driving-related data (e.g., current speed) since such information would increase their feeling of safety ($n = 5$). In contrast, three participants argued that they would need such information only at lower automation levels. P14, P17, and P20 would want the system to explain its (planned) driving actions (e.g., turning or parking). Several participants preferred controls for passenger interaction, e.g., touchscreen- or speech-based input options to customize the visualization display, navigate to a particular destination, or change the route or emergency buttons and functions to contact human support or a (remote) operator.

4.5 Wizard-of-Oz

After lifting the deception and explaining the WoOz setup, 22 of the 30 participants (73 %) stated that they believed that the vehicle was driving autonomously and that the driver was only there for safety reasons. An exploratory analysis revealed a significant correlation between participants doubting the WoOz illusion and their ATI scores (Spearman's rho: $r_s = .411, p = .001$). This indicates that participants with a higher affinity for technology interaction were less likely to believe the deception. However, no other meaningful correlation was found between participants' belief in the autonomous ride and their quantitative assessment of the dependent variables. Thus, we do not differentiate the results based on that. In the following sections, we report detailed findings on the WoOz deception and cover story, participants' driving experience, environmental conditions of the test drives, and the prototype's fidelity.

4.5.1 Deception and Cover Story. Many participants who believed the deception commended the smooth, forward-looking, and defensive driving style (e.g., P2: *"The system mimics an exemplary driver"*; P25: *" When you drive yourself, it's usually not so smooth"*) and were surprised when we lifted the deception (e.g., P3: *"okay, I would have been sure that it drives automated"*). Some comments highlighted the importance of a thoughtful cover story. E.g., P27: *"It was good that you said [the AV] didn't have downtown approval yet, or I probably wouldn't have bought it off"*) and pointed out that the used vehicle's appearance and trust in a certain brand or manufacturer also affect the believability of system capabilities (P13: *"Such a new Mercedes ... that also helped. You tell yourself that it can do nothing wrong."*). However, others regarded the smooth driving style as an indicator that the vehicle could not have been driven by a machine only (e.g., P13: *"from my experience, that was too forward-looking"*). In some situations, that forward-looking driving style was not possible, or the driving wizard failed to conduct it. This led some participants to doubt the autonomous ride (e.g., P30: *"[the ride] was not anticipatory enough for me. So it was two times somehow that the traffic light was yellow and [the vehicle] decided to cancel at short notice"*). Other participants, who doubted the autonomous ride, missed visible sensor hardware indicating that the car is capable of autonomous driving or noticed the wizard's movements (e.g., P4: *"I heard [...] the use of the steering wheel when we were driving "*). P8 explained its doubts with prior knowledge of the current state of technology.

4.5.2 Ride Experience. At the end of the rides, 13 participants commended the positive driving experience (e.g., P12: *"Perfect. Not so abrupt [... but] nice and steady"*; P21: *"it was definitely a very pleasant ride [... and] very interesting"*). Nine participants felt safe because of the safety driver's presence (e.g., P3: *"I had confidence that the safety driver would intervene, if required."*). Four others said they felt safe because of the automated system only. Seven participants compared the ride in the (simulated) AV with being a passive passenger in a taxi or bus. However, some participants had different expectations (e.g., P14: *"I actually imagined autonomous driving to be [...] a softer way of driving"*). While a few participants felt unease due to the video see-through-based WoOz setup (section 4.1.5), others were not bothered by the setup at all (e.g., P22: *"I think that was totally realistic, [...] the image [...] was just fitting to the movements [...] it was [...] as if I was looking out of the front"*). Three participants mentioned that the view through the digital screen affected their perception of the ride (e.g., P11: *"You somehow feel it [...] as a faster ride on the screen than in real life"*).

4.5.3 Environmental Conditions. All test rides were performed during daytime in an urban area with moderate traffic density. Regarding the weather, most of the test rides were conducted under cloudy conditions ($n = 26$). In four sessions, it was rainy, in one snowy, and in 11 sessions, it was (partly) sunny. In the latter, six participants mentioned that the video feed was sometimes overexposed during the ride due to direct sunlight. During the rainy rides, the view out of the vehicle through the windows and consequently through the camera stream was (partly) impaired. However, as there was no heavy rain, the object detection kept functioning correctly.

4.5.4 Prototype Fidelity. Due to technical constraints (hardware and software), the prototype's performance was limited. Some objects were detected late or not at all (mentioned in 12 sessions). In such cases, it was not clear to participants how the visualized objects were selected (variant 2: n = 12; variant 3: n = 9). While we briefed participants that the tested HMI prototype's accuracy is limited due to its early development stage and unlinked from the (simulated) AV's sensors, several participants were disturbed by the offset of the vehicle's driving behavior and the visualization (e.g., P27: *"the car has already braked before the traffic light was even recognized* "). Participants also described some issues and limitations, e.g., occasional image stuttering of the video feed and visualization were considered irritating (n = 6; P28: *"It has quite a bit of flickering every now and then, which is a bit annoying."*;).

5 DISCUSSION

The conducted on-road WoOz study provides design and prototyping learnings for suitable AV HMIs. Furthermore, it provides insights into the potential of transparent information and AV passenger experiences in general. In contrast to previous works, we investigated feedback from an AI-based system with an empirical study in a natural urban environment. The study featured the value of WoOz for context-based prototyping of HMIs that use real-time information and AR. Regarding our research questions and setup, we focus the following discussion on (1) object visualization and (2) passengers' information requirements regarding the tested HMI concepts, pursued by a discourse of (3) the WoOz-based prototype's potentials and considerations, as well as (4) limitations and future work.

5.1 Object Visualization Can Increase Acceptance and UX

With regard to RQ1, qualitative and quantitative results confirm that visual system feedback on detected objects can increase AV acceptance and UX. Concerning the way this information should be presented in the AV windshield (RQ2), most participants preferred the concept with AR overlays (variant 3) over both the baseline concept without information on detected objects (variant 1) and the status bar with object counts only (variant 2). The augmentations were considered helpful in understanding the context better and building trust in the system. This is confirmed by significantly higher understandability and predictability assessments. The results confirm the findings of the online study of Colley et al. [13], who reported increased situation awareness of drivers through (AR-based) visualizations in conditional automated driving. Significantly better evaluations of perceived usefulness and hedonic quality further support the potential of AR-based object visualization, exhibiting an improved UX. Furthermore, many participants reported that object visualizations increased the feeling of safety and trust in the autonomous system.

Although the general positive assessments of the AR overlays, they were in half of the conducted sessions described as too much, irritating, or distracting. This is in line with the findings of Kim et al. [37] on driver distractions induced by AR in vehicles with lower automation levels. The second largest group of study participants considered it sufficient to have only general information on the overall system state, which was the case in the baseline variant. Some participants would not want continuous system feedback at all since they would not want to be distracted from other tasks. The status bar with the object counts served for some participants as an explanation of the AR overlays (variant 3), but was considered not helpful when used alone (variant 2). Some participants did not want to have object visualization at all and favoured variant 1. A reason for this could be negative perceptions of the visualizations, especially when occurring errors degrade the attributed system capabilities (see Section 5.4), which is in line with the findings of another online study by Colley et al. [15]. Furthermore, the rather salient design of the bounding box overlays might have been perceived negatively. Less obtrusive designs (e.g., less salient colors or colored borders only) might be more suitable. Only a few participants would want an intermediate solution, e.g., in the form of the status bar offered by variant 2.

The results generally confirm the potential of transparent system feedback to increase AV acceptance and UX, but not all people would want this information (all the time). Furthermore, the assessed concepts are only early-stage variants. They cover only a small part of the vast possibilities, especially regarding higher-performing hardware and algorithms that might enable more advanced and accurate object visualizations.

5.2 Passengers Want Configurability and Travel Information

Most participants argued for configurable display settings providing passengers with options to select what, when, and how information on the environment and the AVs' reasoning is displayed (RQ2). This aligns with the findings of Oliveira et al. [53], who pointed out that AV passengers might not want contextual information displayed permanently. Participants in our study argued for configurability and on-demand information retrieval (i.e., the HMI should allow them to turn certain information on and off). Regarding the "what", most participants preferred visualizing only objects with an impact on their ride and a visual classification according to their hazard level instead of having permanent visual feedback on detected objects. Future work may investigate respective visualization designs.

The hedonic and pragmatic UX assessment of the three concept variants is in comparison to the UEQ benchmark [60] relatively poor. Qualitative results indicate that this can be (partly) attributed to missing expected travel information, e.g., current location, planned route, and upcoming maneuvers. I.e., since the tested concepts solely focused on providing information on detected objects and system status, participants missed journey-related information. As this was mentioned by almost half of the sample, travel information seems crucial in AVs. This result can be linked to findings of related work on lower automation levels, e.g., [22]. We recommend future work to consider the interplay of novel (information and visualization) concepts with such expected information and to investigate them as a part of holistic interaction concepts.

5.3 Wizard-of-Oz Setup and HMI Prototype Can Serve as Mutual Enablers

The applied WoOz-based setup served two purposes. First, it paved the ground for creating an on-road AV simulation for context-based prototyping and evaluation. Second, it enabled using a computer-vision-based real-time object detection system as an HMI component of a futuristic AR windshield. Nevertheless, WoOz settings also come with methodological challenges [3, 50] that need to be considered, such as keeping up the deception throughout the study. Considering these and the lack of comparable benchmarks, we regard having 73 % of participants believe that the vehicle was driving autonomously until the end of their sessions as evidence of a successful application of the WoOz paradigm. Our exploratory analysis did not reveal statistically relevant correlations of the dependent variables with participants' belief in the WoOz illusion, which is why we did not differentiate the results based on that. Further associated aspects and limitations are discussed in Section 5.4.

In addition to a smooth, defensive, and proactive driving style as recommended by related work [1, 50], we found having suitable hardware with a modern and technologically-advanced appearance (i.e., a vehicle believed to be capable of autonomous driving) as quite supportive of keeping up the deception. Overall, a thoughtful cover story seems to be a crucial part of the WoOz deception. In our case, we consider shifting participants' attention toward the futuristic HMI prototype beneficial. To do this, we told them that we were evaluating new concepts for not yet available hardware components (the AR windshield) and were, thus, requiring the TV-based setup. As a result, WoOz and the windshield interface prototype were mutually beneficial and enabled their successful application.

Nonetheless, not all participants believed the story. Reasons for the doubts can be allocated to, e.g., difficulties in constantly maintaining the defined driving style (e.g., when unexpected events occur), previous knowledge of participants on the state of technology, or observations of participants (e.g., driving-related noises of the wizard). Furthermore, while many participants described the test rides as pleasant, some participants noticed that the video see-through based setup made them feel at unease. This might have been due to the indirect view out of the vehicle and the camera's offset, as well as to the display of the visualizations.

To sum up and answer RQ3, the created WoOz-based prototyping framework served as a suitable basis for this study and may get used and adapted to address similar questions. We recommend future work to thoroughly craft their prototypes, setups, and cover stories and leverage their symbiosis. The prototyping approach can be optimized for future studies according to our descriptions and findings.

5.4 Limitations and Future Work

In the following sections, we discuss the limitations of this paper and the consequent potential for future work regarding (1) the study sample, (2) the HMI prototype, and (3) the applied WoOz approach.

5.4.1 Study Sample. The study sample is characterized by a medium-high affinity for technology interaction. While this is considered a common phenomenon in HCI studies [26], it might impair external validity and affect the belief in the WoOz deception, as the correlation revealed by our exploratory analysis suggests. Furthermore, as is often the case in usability testings, study participants experienced the evaluated system and HMI concepts for the first and only time. However, users' attitudes toward certain aspects can change over time. Future work might conduct long(er)-term studies to account for this circumstance. We also want to note that the study was conducted during the COVID-19 pandemic. Therefore, we applied precautions and hygiene measures (e.g., distancing, wearing medical/FFP2 masks, disinfection of surfaces and hands) and followed local and national authorities' regulations and recommendations. While we consider the pandemic's actual effect on the study results to be minor, it might have affected the sample composition as, e.g., only people without fear of COVID-19 might have signed up for the study in the first place.

5.4.2 HMI Prototype. The evaluated concept variants had a relatively narrow focus on object detection visualization on a (prototyped) windshield interface. We assume that this affected the overall assessment of the rides and visualization concepts. Since the targeted acceptance and UX challenges cannot be addressed in this narrow scope alone, future work should consider the integration with "holistic" HMI concepts (e.g., including visual and auditory passenger information on the planned route and upcoming stops [23]). This would also allow for further investigation of the design space, e.g., in terms of other visualization concepts such as 3D representation in a GUI-based map [65, 70], situation prediction visualization [15], and other feedback modalities such as auditory, kinesthetic, or tactile [33]. Future work may also investigate the HMI configurability suggested by participants and identify relevant situations, maneuvers, objects, or levels of criticality in which information and explanations would be (not) beneficial. As mentioned in Sections 3.1.2 and 4.5.4, the prototype's hardware and performance were limited and consequently not as powerful as cutting-edge sensing systems. This resulted, to some extent, in flickering, missed objects, and classification errors. While the algorithm proved quite robust on rainy rides, extreme lighting conditions (e.g., direct sunlight) resulted in overexposure of the video feed and consequent impairments of sight and object detection. Nevertheless, the used hardware and algorithms served as a suitable basis for the early concept study, the straightforward realization of the AR windshield prototype with real-time information visualization, and the initial investigation of our research questions in an early development phase. Future work may use more powerful industrial hardware and software along with more graphical and computational processing power to enable the use of larger and higher-performing models. Furthermore, adding additional object classes to the model (e.g., animals, construction sites, or hazardous objects) and investigating other visualization approaches might be interesting depending on examined scenarios and conceptual considerations.

5.4.3 Wizard-of-Oz Approach. Since actual AVs are still only available under limited conditions in urban environments, we applied the WoOz paradigm to create a prototyping framework that enabled us to consider the dynamic urban context in our investigation. The approach offers several advantages – especially concerning the evaluated real-time visualization prototype. However, it also poses challenges regarding objectivity, validity, and reliability [50]. While we aimed to control the study as much as possible, particularly dynamic influences cannot get ruled out completely. Furthermore, some of our participants were not fully convinced by the WoOz deception. Since we found no statistically relevant correlation between participants' belief in the WoOz deception with our dependent variables, we did neither exclude data nor create groups based on this. However, we cannot rule out possible effects on the results. Besides the challenges and limitations mentioned, future work may use the described WoOz approach and AR windshield prototype with the reported learnings to conduct further empirical studies. Such studies could investigate AV passenger experiences in a real-world context and HMI concepts relying on real-time information, e.g., visualizations of scene detection, scene prediction, and maneuver planning [15]. Considering the effort to conduct a WoOz study, researchers might, in a first step, formatively evaluate their designs with simpler study designs (e.g., online or simulator studies). In our case, for example, the desire for configurability could have been discovered earlier so that the results could have been incorporated into the subsequent WoOz study. The framework may be further used to prototype AR-based infotainment systems that provide contextual information, e.g., on landmarks or other points of interest [5]. Furthermore, future work could focus on the method itself and investigate the effect of participants' belief in the WoOz deception, e.g., by comparing one group that is truthfully informed about the system's actual capabilities with another group that gets told the WoOz cover story.

6 CONCLUSIONS

Suitable HMI concepts are required to address AVs' acceptance and UX challenges. The conducted on-road WoOz study with 30 participants evaluating early visualization concepts for a windshield interface confirms the potential of transparent communication and object detection visualization to increase the acceptance and UX of AVs.

System feedback on detected objects was deemed useful, and AR-based visualization, in particular, significantly increased the system's understandability and predictability, perceived usefulness, and hedonic quality. However, in line with related work from online surveys, lab studies, and other automation levels, we found that (permanent) system feedback can also annoy, irritate, or distract passengers. We identified making the information configurable for individual user requirements and accessible on-demand as a promising approach to address this challenge. In addition, as travel-related information (e.g., current location, planned route, and upcoming stops) is essential in driverless vehicles, it needs to be investigated how transparent system feedback can be integrated with such information when designing holistic AV HMI concepts.

The applied video-based WoOz approach provides a suitable framework for prototyping both AVs and (AR-based) windshield interfaces with real-time information visualization. However, it poses technological and methodological challenges. A compelling cover story is essential for keeping up the WoOz deception and the study's success. It can be supported by fitting hardware (e.g., a modern vehicle) and an appropriate "AV-like" driving style.

To sum up, this work contributes to the human-centered design of human-AV interactions. It demonstrates a straightforward WoOz-based method for context-based prototyping of (AR-based) real-time AV HMIs that is suggested to be adopted and advanced by future work. Furthermore, it provides learnings and practical recommendations for system design and future studies.

ACKNOWLEDGMENTS

This research has partly received funding from the German Federal Ministry of Transport and Digital Infrastructure (BMVI) within the funding guideline "Automated and Connected Driving" under grant number 16AVF2134G and from the German Federal Ministry of Economic Affairs and Climate Action (BMWK) within the funding guideline "New Vehicle and System Technologies" under grant number 19A22006K. We want to thank our colleagues at Ergosign for their support and valuable feedback in the study's preparation, conduct, and analysis. Particularly, we thank Michael Alexander for his great ideas and on-point support in the mechanical study setup. Furthermore, we thank the study participants for their participation and the anonymous reviewers for their helpful feedback.

REFERENCES

[1] Sonia Baltodano, Srinath Sibi, Nikolas Martelaro, Nikhil Gowda, and Wendy Ju. 2015. The RRADS Platform: A Real Road Autonomous Driving Simulator. In *Proceedings of the 7th International Conference on Automotive User Interfaces and Interactive Vehicular Applications* (Nottingham, United Kingdom) *(AutomotiveUI '15)*. Association for Computing Machinery, New York, NY, USA, 281–288. https://doi.org/10.1145/2799250.2799288

[2] Klaus Bengler, Klaus Dietmayer, Berthold Färber, Markus Maurer, Christoph Stiller, and Hermann Winner. 2014. Three Decades of Driver Assistance Systems – Review and Future Perspectives. *IEEE Intelligent Transportation Systems Magazine* 6, 4 (2014), 6–22. https://doi.org/10.1109/MITS.2014.2336271

[3] Klaus Bengler, Kamil Omozik, and Andrea Isabell Müller. 2019. The Renaissance of Wizard of Oz (WoOz) - Using the WoOz Methodology to Prototype Automated Vehicles. In *Proceedings of the Human Factors and Ergonomics Society Europe Chapter 2019 Annual Conference*. 63–72. https://www.researchgate.net/publication/346659448

[4] Klaus Bengler, Michael Rettenmaier, Nicole Fritz, and Alexander Feierle. 2020. From HMI to HMIs: Towards an HMI Framework for Automated Driving. *Information* 11, 2 (2020). https://doi.org/10.3390/info11020061

[5] Melanie Berger, Aditya Dandekar, Regina Bernhaupt, and Bastian Pfleging. 2021. An AR-Enabled Interactive Car Door to Extend In-Car Infotainment Systems for Rear Seat Passengers. In *Extended Abstracts of the 2021 CHI Conference on Human Factors in Computing Systems* (Yokohama, Japan) *(CHI EA '21)*. Association for Computing Machinery, New York, NY, USA, Article 404, 6 pages. https://doi.org/10.1145/3411763.3451589

[6] Niels Ole Bernsen, Hans Dybkjær, and Laila Dybkjær. 1993. Wizard of Oz Prototyping: When and How? *Cognitive Science* 94 (1993), 1–13. http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.135.2044&rep=rep1&type=pdf

[7] Alexey Bochkovskiy, Chien-Yao Wang, and Hong-Yuan Mark Liao. 2020. YOLOv4: Optimal Speed and Accuracy of Object Detection. (4 2020). https://doi.org/10.48550/arXiv.2004.10934

[8] Richard E. Boyatzis. 1998. *Transforming Qualitative Information: Thematic Analysis and Code Development.* Sage, Thousand Oaks, CA.

[9] Virginia Braun and Victoria Clarke. 2006. Using thematic analysis in psychology. *Qualitative Research in Psychology* 3, 2 (2006), 77–101. https://doi.org/10.1191/1478088706qp063oa

[10] Raluca Budiu. 2018. Between-Subjects vs. Within-Subjects Study Design. https://www.nngroup.com/articles/between-within-subjects/

[11] Ching-Fu Chen. 2019. Factors affecting the decision to use autonomous shuttle services: Evidence from a scooter-dominant urban context. *Transportation Research Part F: Traffic Psychology and Behaviour* 67 (2019), 195–204. https://doi.org/10.1016/j.trf.2019.10.016

[12] Jacob Cohen. 1992. A Power Primer. *Psychological Bulletin* 112, 1 (1992), 155–159. https://doi.org/10.1037/0033-2909.112.1.155

[13] Mark Colley, Benjamin Eder, Jan Ole Rixen, and Enrico Rukzio. 2021. Effects of Semantic Segmentation Visualization on Trust, Situation Awareness, and Cognitive Load in Highly Automated Vehicles. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems* (Yokohama, Japan) *(CHI '21)*. Association for Computing Machinery, New York, NY, USA, Article 155, 11 pages. https://doi.org/10.1145/3411764.3445351

[14] Mark Colley, Svenja Krauss, Mirjam Lanzer, and Enrico Rukzio. 2021. How Should Automated Vehicles Communicate Critical Situations? A Comparative Analysis of Visualization Concepts. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 5, 3, Article 94 (sep 2021), 23 pages. https://doi.org/10.1145/3478111

[15] Mark Colley, Max Rädler, Jonas Glimmann, and Enrico Rukzio. 2022. Effects of Scene Detection, Scene Prediction, and Maneuver Planning Visualizations on Trust, Situation Awareness, and Cognitive Load in Highly Automated Vehicles. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 6, 2, Article 49 (jul 2022), 21 pages. https://doi.org/10.1145/3534609

[16] Condens Insights GmbH. 2022. Supercharge your UX research analysis with condens. https://condens.io/

[17] John W. Creswell. 2014. *A Concise Introduction to Mixed Methods Research.* Sage Publications, Thousand Oaks, California, USA.

[18] Fred D. Davis. 1986. *A Technology Acceptance Model for Empirically Testing New End-User Information Systems: Theory and Results.* Thesis (Ph. D.). Massachusetts Institute of Technology. https://doi.org/1721.1/15192

[19] Joost C.F. de Winter, Peter M. van Leeuwen, and Riender Happee. 2012. Advantages and Disadvantages of Driving Simulators: A Discussion. In *Proceedings of the 8th International Conference on Methods and Techniques in Behavioral Research (Measuring Behavior '12)*, Andrew Spink, Fabrizio Grieco, Olga Krips, Leanne Loijens, Lucas Noldus, and Patrick Zimmerman (Eds.). Noldus Information Technology, Utrecht„ 47–50.

[20] Henrik Detjen, Bastian Pfleging, and Stefan Schneegass. 2020. A Wizard of Oz Field Study to Understand Non-Driving-Related Activities, Trust, and Acceptance of Automated Vehicles. In *12th International Conference on Automotive User Interfaces and Interactive Vehicular Applications* (Virtual Event, DC, USA) *(AutomotiveUI '20)*. Association for Computing Machinery, New York, NY, USA, 19–29. https://doi.org/10.1145/3409120.3410662

[21] Grace Eden, Benjamin Nanchen, Randolf Ramseyer, and Florian Evèquoz. 2017. On the road with an autonomous passenger shuttle: Integration in public spaces. *Conference on Human Factors in Computing Systems - Proceedings* Part F1276, October (2017), 1569–1576. https://doi.org/10.1145/3027063.3053126

[22] Alexander Feierle, Simon Danner, Sarah Steininger, and Klaus Bengler. 2020. Information needs and visual attention during urban, highly automated driving-An investigation of potential influencing factors. *Information* 11, 2 (2020), 1–17. https://doi.org/10.3390/info11020062

[23] Lukas A. Flohr, Dominik Janetzko, Dieter P. Wallach, Sebastian C. Scholz, and Antonio Krüger. 2020. Context-Based Interface Prototyping and Evaluation for (Shared) Autonomous Vehicles Using a Lightweight Immersive Video-Based Simulator. In *Proceedings of the 2020 ACM Designing Interactive Systems Conference* (Eindhoven, Netherlands) *(DIS '20)*. Association for Computing Machinery, New York, NY, USA, 1379–1390. https://doi.org/10.1145/3357236.3395468

[24] Lukas A. Flohr, Sofie Kalinke, Antonio Krüger, and Dieter P. Wallach. 2021. Chat or Tap? – Comparing Chatbots with 'Classic' Graphical User Interfaces for Mobile Interaction with Autonomous Mobility-on-Demand Systems. In *Proceedings of the 23rd International Conference on Mobile Human-Computer Interaction* (Toulouse & Virtual, France) *(MobileHCI '21)*. Association for Computing Machinery, New York, NY, USA, Article 21, 13 pages. https://doi.org/10.1145/3447526.3472036

[25] Lukas A. Flohr and Dieter P. Wallach. 2023. The Value of Context-Based Interface Prototyping for the Autonomous Vehicle Domain: A Method Overview. *Multimodal Technologies and Interaction* 7, 1 (2023), 1–17. https://doi.org/10.3390/mti7010004

[26] Thomas Franke, Christiane Attig, and Daniel Wessel. 2019. A Personal Resource for Technology Interaction: Development and Validation of the Affinity for Technology Interaction (ATI) Scale. *International Journal of Human-Computer Interaction* 35, 6 (2019), 456–467. https://doi.org/10.1080/10447318.2018.1456150

[27] Michael A. Gerber, Ronald Schroeter, and Julia Vehns. 2019. A Video-Based Automated Driving Simulator for Automotive UI Prototyping, UX and Behaviour Research. In *Proceedings of the 11th International Conference on Automotive User Interfaces and Interactive Vehicular Applications* (Utrecht, Netherlands) *(AutomotiveUI '19)*. Association for Computing Machinery, New York, NY, USA, 14–23. https://doi.org/10.1145/3342197.3344533

[28] German Research Center for Artificial Intelligence (DFKI). 2019. OpenDS: Open Source Driving Simulation. https://opends.dfki.de/

[29] Andrew J. Hawkins. 2019. Waymo's self-driving cars are now available on Lyft's app in Phoenix - The Verge. https://www.theverge.com/2019/5/7/18536003/waymo-lyft-self-driving-ride-hail-app-phoenix

[30] Ann-Christin Hensch, Isabel Neumann, Matthias Beggiato, Josephine Halama, and Josef F. Krems. 2019. Effects of a light-based communication approach as an external HMI for Automated Vehicles - A Wizard-of-Oz Study. *Transactions on Transport Sciences* 10, 2 (2019), 18–32. https://doi.org/10.5507/TOTS.2019.012

[31] Hamburger Hochbahn. 2021. The HEAT project. https://www.hochbahn.de/en/projects/the-heat-project

[32] Marius Hoggenmüller, Martin Tomitsch, Luke Hespanhol, Tram Thi Minh Tran, Stewart Worrall, and Eduardo Nebot. 2021. Context-Based Interface Prototyping: Understanding the Effect of Prototype Representation on User Feedback. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems* (Yokohama, Japan) *(CHI '21)*. Association for Computing Machinery, New York, NY, USA, Article 370, 14 pages. https://doi.org/10.1145/3411764.3445159

[33] Pascal Jansen, Mark Colley, and Enrico Rukzio. 2022. A Design Space for Human Sensor and Actuator Focused In-Vehicle Interaction Based on a Systematic Literature Review. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 6, 2, Article 56 (jul 2022), 51 pages. https://doi.org/10.1145/3534617

[34] JASP Team. 2022. JASP (Version 0.16.1)[Computer software]. https://jasp-stats.org/

[35] Juffrizal Karjanto, Nidzamuddin Md. Yusof, Jacques Terken, Frank Delbressine, Matthias Rauterberg, and Muhammad Zahir Hassan. 2018. Development of On-Road Automated Vehicle Simulator for Motion Sickness Studies. *International Journal of Driving Science* 1, 1 (2018), 1–12. https://doi.org/10.5334/ijds.8

[36] Kanwaldeep Kaur and Giselle Rampersad. 2018. Trust in driverless cars: Investigating key factors influencing the adoption of driverless cars. *Journal of Engineering and Technology Management* 48 (2018), 87–96. https://doi.org/10.1016/j.jengtecman.2018.04.006

[37] Hyungil Kim and Joseph L. Gabbard. 2022. Assessing Distraction Potential of Augmented Reality Head-Up Displays for Vehicle Drivers. *Human Factors* 64 (8 2022), 852–865. Issue 5. https://doi.org/10.1177/0018720819844845

[38] Keunwoo Kim, Minjung Park, and Youn-kyung Lim. 2021. Guiding Preferred Driving Style Using Voice in Autonomous Vehicles: An On-Road Wizard-of-Oz Study. In *Designing Interactive Systems Conference 2021* (Virtual Event, USA) *(DIS '21)*. Association for Computing Machinery, New York, NY, USA, 352–364. https://doi.org/10.1145/3461778.3462056

[39] Sangwon Kim, Jennifer Jah Eun Chang, Hyun Ho Park, Seon Uk Song, Chang Bae Cha, Ji Won Kim, and Namwoo Kang. 2020. Autonomous Taxi Service Design and User Experience. *International Journal of Human-Computer Interaction* 36, 5 (2020), 429–448. https://doi.org/10.1080/10447318.2019.1653556

[40] René F. Kizilcec. 2016. How Much Information? Effects of Transparency on Trust in an Algorithmic Interface. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems* (San Jose, California, USA) *(CHI '16)*. Association for Computing Machinery, New York, NY, USA, 2390–2395. https://doi.org/10.1145/2858036.2858402

[41] Moritz Körber. 2019. Theoretical considerations and development of a questionnaire to measure trust in automation. In *Proceedings of the 20th Congress of the International Ergonomics Association (IEA 2018): Volume VI: Transport Ergonomics and Human Factors (TEHF), Aerospace Human Factors and Ergonomics* (1 ed.). 13–30. http://link.springer.com/10.1007/978-3-319-96074-6_2

[42] Andreas Korthauer, Clemens Guenther, Andreas Hinrichs, Wen Ren, and Yiwen Yang. 2021. Watch Your Vehicle Driving at the City: Interior HMI with Augmented Reality for Automated Driving. In *22nd International Conference on Human-Computer Interaction with Mobile Devices and Services* (Oldenburg, Germany) *(MobileHCI '20)*. Association for Computing Machinery, New York, NY, USA, Article 59, 5 pages. https://doi.org/10.1145/3406324.3425895

[43] Christian Kray, Patrick Olivier, Amy Weihong Guo, Pushpendra Singh, Hai Nam Ha, and Phil Blythe. 2007. Taming context: A key challenge in evaluating the usability of ubiquitous systems. In *Ubiquitous Systems Evaluation 2007 (USE '07) - Workshop at Ubicomp 2007.*

[44] Sven Krome, William Goddard, Stefan Greuter, Steffen P. Walz, and Ansgar Gerlicher. 2015. A Context-Based Design Process for Future Use Cases of Autonomous Driving: Prototyping AutoGym. In *Proceedings of the 7th International Conference on Automotive User Interfaces and Interactive Vehicular Applications* (Nottingham, United Kingdom) *(AutomotiveUI '15)*. Association for Computing Machinery, New York, NY, USA, 265–272. https://doi.org/10.1145/2799250.2799257

[45] Andrew R. Lacher, Robert Grabowski, and Stephen Cook. 2014. Autonomy, Trust, and Transportation. In *Proceedings of the 2014 AAAI Spring Symposium.* 42–49.

[46] Tsung-Yi Lin, Michael Maire, Belongie Serge, Hays James, Perona Pietro, Ramanan Deva, Dollár Piotrand, and Zitnick C Lawrence. 2014. Microsoft COCO: Common Objects in Context. In *Proceedings of the 2014 European Conference on Computer Vision (ECCV '14)* (Cham), Fleet David, Tomas Pajdla, Schiele Bernt, and Tuytelaars Tinne (Eds.). Springer International Publishing, 740–755. https://doi.org/10.1007/978-3-319-10602-1_48

[47] Patrick Lindemann, Tae Young Lee, and Gerhard Rigoll. 2018. Catch my drift: Elevating situation awareness for highly automated driving with an explanatory windshield display user interface. *Multimodal Technologies and Interaction* 2, 4 (2018). https://doi.org/10.3390/mti2040071

[48] Ana Mackay, Inês Fortes, Catarina Santos, Dário Machado, Patrícia Barbosa, Vera Vilas Boas, João Pedro Ferreira, Nélson Costa, Carlos Silva, and Emanuel Sousa. 2019. The impact of autonomous vehicles' active feedback on trust. In *International Conference on Applied Human Factors and Ergonomics (AHFE 2019): Advances in Safety Management and Human Factors.* 342–352. https://doi.org/10.1007/978-3-030-20497-6_32

[49] Johanna Meurer, Christina Pakusch, Gunnar Stevens, Dave Randall, and Volker Wulf. 2020. A Wizard of Oz Study on Passengers' Experiences of a Robo-Taxi Service in Real-Life Settings. In *Proceedings of the 2020 ACM Designing Interactive Systems Conference* (Eindhoven, Netherlands) *(DIS '20)*. Association for Computing Machinery, New York, NY, USA, 1365–1377. https://doi.org/10.1145/3357236.3395465

[50] Andrea Isabell Müller, Veronika Weinbeer, and Klaus Bengler. 2019. Using the Wizard of Oz Paradigm to Prototype Automated Vehicles: Methodological Challenges. In *Proceedings of the 11th International Conference on Automotive User Interfaces and Interactive Vehicular Applications: Adjunct Proceedings* (Utrecht, Netherlands) *(AutomotiveUI '19)*. Association for Computing Machinery, New York, NY, USA, 181–186. https://doi.org/10.1145/3349263.3351526

[51] Sina Nordhoff, Joost de Winter, Ruth Madigan, Natasha Merat, Bart van Arem, and Riender Happee. 2018. User acceptance of automated shuttles in Berlin-Schöneberg: A questionnaire study. *Transportation Research Part F: Traffic Psychology and Behaviour* 58, October (2018), 843–854. https://doi.org/10.1016/j.trf.2018.06.024

[52] Waypoint The official Waymo blog. 2022. Wheels up for Waymo as we expand our 24/7 rider-only territories. https://blog.waymo.com/2022/12/wheels-up-for-waymo-as-we-expand.html

[53] Luis Oliveira, Christopher Burns, Jacob Luton, Sumeet Iyer, and Stewart Birrell. 2020. The influence of system transparency on trust: Evaluating interfaces in a highly automated vehicle. *Transportation Research Part F: Traffic Psychology and Behaviour* 72 (2020), 280–296. https://doi.org/10.1016/j.trf.2020.06.001

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

[55] Caroline Pigeon, Aline Alauzet, and Laurence Paire-Ficout. 2021. Factors of acceptability, acceptance and usage for non-rail autonomous public transport vehicles: a systematic literature review. *Transportation Research Part F: Traffic Psychology and Behaviour* 81, August (2021), 251–270. https://doi.org/10.1016/j.trf.2021.06.008

[56] Mark O. Riedl. 2019. Human-centered artificial intelligence and machine learning. *Human Behavior and Emerging Technologies* 1, 1 (2019), 33–36. https://doi.org/10.1002/hbe2.117 arXiv:1901.11184

[57] José Rodríguez. 2022. Cruise Self-Driving Taxis Start Daytime Rides in San Francisco. https://jalopnik.com/cruise-self-driving-taxis-start-daytime-rides-in-san-fr-1849798420

[58] SAE International and ISO. 2021. J3016: Surface Vehicle Recommended Practice: Taxonomy and Definitions for Terms Related to Driving Automation Systems for On-Road Motor Vehicles.

[59] Hans-Peter Schöner and Bernhard Morys. 2015. Dynamische Fahrsimulatoren. In *Handbuch Fahrerassistenzsysteme* (3 ed.), H. Winner, S. Hakuli, F. Lotz, and C. Singer (Eds.). Springer Fachmedien, Wiesbaden, Chapter 9, 139–155. https://doi.org/10.1007/978-3-658-05734-3_9

[60] Martin Schrepp, Andreas Hinderks, and Jörg Thomaschewski. 2017. Construction of a Benchmark for the User Experience Questionnaire (UEQ). *International Journal of Interactive Multimedia and Artificial Intelligence* 4, 4 (2017), 40. https://doi.org/10.9781/ijimai.2017.445

[61] Martin Schrepp, Andreas Hinderks, and Jörg Thomaschewski. 2017. Design and Evaluation of a Short Version of the User Experience Questionnaire (UEQ-S). *International Journal of Interactive Multimedia and Artificial Intelligence* 4, 6 (2017), 103. https://doi.org/10.9781/ijimai.2017.09.001

[62] Martina Schuß, Philipp Wintersberger, and Andreas Riener. 2021. Let's Share a Ride into the Future: A Qualitative Study Comparing Hypothetical Implementation Scenarios of Automated Vehicles. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems* (Yokohama, Japan) *(CHI '21)*. Association for Computing Machinery, New York, NY, USA, Article 156, 11 pages. https://doi.org/10.1145/3411764.3445609

[63] Ben Shneiderman. 2020. Human-Centered Artificial Intelligence: Three Fresh Ideas. *AIS Transactions on Human-Computer Interaction* 12, 3 (2020), 109–124. https://doi.org/10.17705/1thci.00131

[64] Kevin Spieser, Kyle Treleaven, Rick Zhang, Emilio Frazzoli, Daniel Morton, and Marco Pavone. 2014. Toward a Systematic Approach to the Design and Evaluation of Automated Mobility-on-Demand Systems: A Case Study in Singapore. In *Road Vehicle Automation*, Gereon Meyer and Sven Beiker (Eds.). Springer International Publishing, Basel, 229–245. https://doi.org/10.1007/978-3-319-05990-7_20

[65] Tesla. 2022. Autopilot | Model 3 and Model Y. https://www.youtube.com/watch?v=-FeMwPUAOLM

[66] The jamovi Project. 2022. jamovi. https://www.jamovi.org/

[67] Nikolas Thomopoulos and Moshe Givoni. 2015. The autonomous car—a blessing or a curse for the future of low carbon mobility? An exploration of likely vs. desirable outcomes. *European Journal of Futures Research* 3, 14 (2015), 14. https://doi.org/10.1007/s40309-015-0071-z

[68] Jinke D. Van der Laan, Adriaan Heino, and Dick De Waard. 1997. A simple procedure for the assessment of acceptance of advanced transport telematics. *Transportation Research Part C: Emerging Technologies* 5, 1 (1997), 1–10. https://doi.org/10.1016/S0968-090X(96)00025-3

[69] Peter Wang, Srinath Sibi, Brian Mok, and Wendy Ju. 2017. Marionette: Enabling On-Road Wizard-of-Oz Autonomous Driving Studies. *ACM/IEEE International Conference on Human-Robot Interaction* Part F127194 (2017), 234–243. https://doi.org/10.1145/2909824.3020256

[70] Waymo. 2022. Taking a rider-only trip with the Waymo Driver in San Francisco. https://www.youtube.com/watch?v=ospoTAyEdDQ

[71] Daniel Wessel, Christiane Attig, and Thomas Franke. 2019. ATI-S - An Ultra-Short Scale for Assessing Affinity for Technology Interaction in User Studies. In *Proceedings of Mensch Und Computer 2019* (Hamburg, Germany) *(MuC'19)*. Association for Computing Machinery, New York, NY, USA, 147–154. https://doi.org/10.1145/3340764.3340766

[72] Philipp Wintersberger, Anna-Katharina Frison, Andreas Riener, and Tamara von Sawitzky. 2018. Fostering User Acceptance and Trust in Fully Automated Vehicles: Evaluating the Potential of Augmented Reality. *Presence: Teleoperators and Virtual Environments* 27, 1 (02 2018), 46–62. https://doi.org/10.1162/pres_a_00320

[73] Philipp Wintersberger, Tamara von Sawitzky, Anna-Katharina Frison, and Andreas Riener. 2017. Traffic Augmentation as a Means to Increase Trust in Automated Driving Systems. In *Proceedings of the 12th Biannual Conference on Italian SIGCHI Chapter* (Cagliari, Italy) *(CHItaly '17)*. Association for Computing Machinery, New York, NY, USA, Article 17, 7 pages. https://doi.org/10.1145/3125571.3125600

[74] World Medical Association. 2000. World Medical Association Declaration of Helsinki – Ethical Principles for Medical Research Involving Human Subjects.

[75] Larissa Zacherl, Jonas Radlmayr, and Klaus Bengler. 2020. Constructing a Mental Model of Automation Levels in the Area of Vehicle Guidance. In *Proceedings of the 3rd International Conference on Intelligent Human Systems Integration (IHSI '20): Integrating People and Intelligent Systems*. Springer International Publishing, Modena, Italy, 73–79. https://doi.org/10.1007/978-3-030-39512-4_12