Chat or Tap? – Comparing Chatbots with 'Classic' Graphical User Interfaces for Mobile Interaction with Autonomous Mobility-on-Demand Systems

LUKAS A. FLOHR, Ergosign GmbH, Germany and Saarland Informatics Campus, Germany

SOFIE KALINKE, Ergosign GmbH, Germany

ANTONIO KRÜGER, DFKI GmbH, Germany and Saarland Informatics Campus, Germany

DIETER P. WALLACH, Ergosign GmbH, Germany

In autonomous mobility-on-demand (AMoD) systems, passengers will solely interact with autonomous vehicles via digital user interfaces (UIs). Hence, UIs are crucial for acceptance and user experience (UX). As a foundation for deriving empirically grounded design guidelines, we investigate two approaches for mobile interaction: chatbots and 'classic' graphical UIs (GUIs). We evaluated prototypes of both in expert studies ($n_{GUI} = 6$; $n_{Chatbot} = 5$) and a between-subjects simulator user study ($n = 34$). The latter enabled participants to experience a complete AMoD journey. While both concepts receive good acceptance and positive UX evaluations, the GUI results in significantly higher attractiveness and user satisfaction ratings. A significant interaction effect reveals a higher intention to use the chatbot in scenarios with a change of plans, but the GUI in 'happy path' scenarios. Interview data and emotion curves support this effect. Balancing the concepts' advantages and disadvantages, we provide design recommendations and propose to use GUI-based mobile applications with integrated (text-based) conversational elements for future human-AMoD interaction.

CCS Concepts: • **Human-centered computing → User studies**; **Natural language interfaces**; **Graphical user interfaces**; **Personal digital assistants**; *Empirical studies in HCI*; *Empirical studies in ubiquitous and mobile computing*.

Additional Key Words and Phrases: Chatbots; Virtual personal assistants; Conversational user interfaces; Graphical user interfaces; Autonomous mobility-on-demand; Ride-sharing; Context-based prototyping; Immersive video-based driving simulation.

ACM Reference Format:

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 *MobileHCI '21: ACM International Conference on Mobile Human-Computer Interaction, Sept. 27 – Oct. 1, 2021, Toulouse, France & virtually*. ACM, New York, NY, USA, 20 pages. https://doi.org/10.1145/1122445.1122456

1 INTRODUCTION

Autonomous mobility-on-demand (AMoD) is a futuristic mode of transport in which passengers are transported via autonomous vehicles (AVs) in a certain environment [37]. Such driverless AVs, classifiable by SAE automation levels 4 and 5 [40, 55], are expected to become a commercial mode of transportation within the next years (e.g., [2]) and will enable the creation of convenient and efficient AMoD systems. Simulations suggest that shared AMoD can meet personal transportation needs in metropolitan cities with only one-third of the current operating passenger vehicles [48]. This expectation illustrates the technology's tremendous economic, societal, and ecological potential.

Comparable to taking a ride in a (shared) taxi, journeys in AMoD are temporally and spatially flexible, i.e., neither fixed timetables nor fixed pick-up or drop-off locations are required. However, given that there will not be a driver or another accompanying human assistant available within the AV (e.g., to answer traveler queries), AMoD varies greatly from current mobility-on-demand services. The new situation of riding in a driverless vehicle might feel awkward to passengers exposed to an autonomous system's decisions and actions. Thus, digital user interfaces (UIs) capable of filling the resulting service and information gap are needed. Such UIs require to compensate for the absence of a human driver and to gain users' trust and public acceptance — which is, besides technological hurdles, a main challenge of future AMoD systems [26].

To counteract acceptance challenges, HCI research needs to provide answers to questions including *how should these intelligent, AI-powered AMoD systems communicate with the user?* and *how can we design and (formatively) evaluate UIs while considering the whole UX – before, during and after the ride?* (e.g., [53]).

Most current HCI research in the field of AMoD focuses primarily on general acceptance aspects of autonomous vehicles (e.g., [3, 6, 35]). Very few studies investigate the actual interaction, i.e., the design and evaluation of UIs (e.g., [13, 27]). However, at the latest when the technology is ready for market introduction, the HCI community needs to be able to provide future practitioners, manufacturers and service providers with guidelines and recommendations for human-centered design and development.

This paper explores concrete approaches for UI design and evaluation to tackle potential acceptance and UX hurdles of AMoD systems from an early development stage. As we are considering the whole user journey, we focus on mobile applications capable of accompanying users through all usage situations and confidentially providing individualized information – also during shared AMoD rides.

Currently, 'classic' GUIs with touchscreen-based interaction can be considered as the status quo for interacting with mobility applications on personal devices. However, AI-infused systems, comprising assistive dialog-based interaction concepts also promise to provide a good or even better fit when it comes to acceptance of new technology. Such support can be accessed via conversational UIs (CUIs) like chatbots (text-based) and virtual assistants (speech-based). CUIs enable people, similar to talking or chatting with a real person, "to interact with smart devices using spoken language in a natural way" [30, p. 1].

The question arises whether CUIs are as good as or even better suited than 'classic' GUIs for interacting with AMoD systems on mobile devices. Especially when it comes to establishing a sensitive communication between users and systems in new usage situations, CUIs seem to provide a promising approach (e.g., [51]). Depending on the situation, the two approaches might also get combined meaningfully. For instance, in many modern cars, drivers can either tap on a touchscreen or talk to a virtual assistant while driving. However, considering shared rides (i.e., multi-user scenarios) and users' privacy requirements (e.g., [4]), speech-based interaction with virtual assistants does not seem to be a good basis for interacting with AMoD systems on personal devices. Instead, classic GUIs and chatbots appear favorable. While GUIs and chatbots are intensively investigated across domains, literature lacks comparisons of the two approaches. If available at all, the transferability of existing comparisons (e.g., [51]) to the AMoD domain is limited. This work aims to fill this research gap and, consequently, investigates the following research question: *are 'classic' GUIs or chatbots better suited to provide the basis for mobile human-AMoD interaction in terms of UX and user acceptance?*

The papers' contribution to the field of mobile HCI is twofold. First, it illustrates and compares the strengths and weaknesses of chatbots and 'classic' GUIs – in general and specifically for the AMoD domain – based on existing literature, expert evaluations, and a comparative user study. Second, it demonstrates how prototypes of intelligent UIs for AMoD systems can be evaluated and holistically compared in early development phases while considering the complete user journey in a video-based simulation environment.

2 RELATED WORK

Besides achieving technological maturity, AMoD poses enormous challenges for HCI. The following sections provide a comprehensive understanding of acceptance and design challenges, AMoD UIs, and context-based evaluation.

2.1 Acceptance and Design Challenges

Viewed from an HCI angle, one of the biggest challenges for AMoD systems are posed by public acceptance [26]. Critical acceptance barriers can be allocated to (adverse) user expectations, (un-)reliability of the underlying technology, overall system performance and security issues, as well as to privacy concerns, and – most importantly – trust issues [26]. To face these challenges, a clear understanding of the interdependence between people, system, and environment is required [28].

Based on this understanding, user interfaces can be designed and developed to adequately substitute aspects of the former driver's role. Future interfaces need to supply users with sufficient situational awareness, enable them to understand actions and signals of the AV and provide them with appropriate functionalities to tell the system and the AV about their intentions and needs [14, 53]. Since flexibility is regarded as one of the main reasons for users to switch from conventional private transport to autonomous mobility services [38], a prevalent challenge for interacting with AVs – in contrast to conventional taxis – is posed by users' spontaneous change of plans (e.g., route changes or stop requests) [36]. Therefore, AMoD UIs also need to provide options to handle such occurring user requests.

Not surprisingly, there is a vivid research stream on (1) how to create acceptable systems providing a trustful and enjoyable user experience and (2) on how to evaluate these from early development phases on with regard to their potential to counteract acceptance barriers (e.g., [15, 24, 53]).

2.2 User Interfaces for AMoD

UIs provide information and controls to accomplish specific tasks with an interactive system [11]. AMoD UIs can range from personal planning and booking applications on various devices (e.g., smartphones, tablets, desktops) to in-vehicle information displays and terminals at mobility hubs with diverse in- and output modalities. For interacting with AVs, users seem to prefer established technologies such as smartphone booking apps, in-vehicle touchscreens and control buttons but tend to reject less familiar methods, including analogue hand gesture communication [3].

Since there is no human driver onboard a (shared) AV, the question arises on how to communicate with the AV and the service throughout the journey – and especially during the ride – considering acceptance, privacy, and trust issues (e.g., [4, 26]). This question is inevitable with regard to occurring change of plans (CoP), e.g., the need to change the departure time or target destination of a booked trip or abort an ongoing trip. In general, the services of a human driver might be substituted by a mobile app [3] serving as a personal travel companion. Such an app could rely on a 'classic' GUI (e.g., [13]) but also on a CUI (e.g., [27]).

As cost-effective AMoD rides would be shared with other passengers, the interaction between users and system will include interacting in public spaces with other people present. Given that travel information can be a private matter [4], visual in- and output seems superior to other interaction modalities like, e.g., speech. As a consequence, 'classic' GUIs and chatbots (i.e., text-based CUIs) appear to be promising approaches for a mobile AMoD companion app. The following sections provide an overview of the two concepts and their advantages and disadvantages discussed in the literature.

2.2.1 *GUIs.*

GUIs with touch-input can be considered as the de facto status quo of interacting with computer systems on mobile devices. This also applies for currently available ride-sharing and mobility-on-demand services like, e.g., Uber, MOIA, CleverShuttle, or Free Now.

Designing for GUIs' usability and positive user experiences implies to think about navigation patterns, menu structures, and the interaction with graphical elements, editable text fields or buttons [11, 16]. Generally, GUIs can – in contrast to CUIs – easily provide an overview of a system's functionalities and scope [32, 49] and are suitable to display plenty of information [5]. Based on that, users can build a clear mental model of the system (e.g., [32]). This makes it easy for them to choose between provided options and discover the system through visual clues (e.g., [32]).

As a result, GUIs can provide efficient shortcuts [50] to access specific functions (e.g., aborting an ongoing (shared) AMoD ride with a single button). Although GUIs often use established concepts, users must be able to understand the layout and visuals, their underlying logic, and the interaction concept [49, 50]. This is a major design challenge, especially concerning new usage contexts like AMoD that come with new functions and restrictions.

2.2.2 Chatbots. In general, "the front-end to a chatbot or virtual personal assistant" is provided by a CUI [31, p. 40] enabling users to interact with natural language and in- and output modalities like, e.g., speech, text, or touch [30, 31]. When users express their needs in their own words [49], the system needs to understand their intents [51].

In contrast to voice-based virtual assistants, chatbots are text-based and require a graphical counterpart. Their visual UI can be considered as a 'blank canvas' providing content and features on demand [16]. Designing for chatbot usability implies providing users with the appropriate information at the right time and making good suggestions [16]. If done right, chatbots can simplify the information search process [51], especially in complex search spaces. Through the similarity with natural conversations and instant messaging apps, they can provide convenience and ease of use [51].

By taking contextual information acquired in previous conversations into account, bots can also personalize the interaction based on individual users' characteristics [51, 56] and offer context-based 'shortcuts' to step-by-step approaches commonly used by 'classic' GUIs. For example, in shared AMoD rides, it might typically not be possible to change the destination of an ongoing trip as this would also affect all other passengers' rides. However, suppose a user requests such a change of plans. In that case, a chatbot could consider the conflict with the ongoing ride, inform the user about it and directly suggest a solution (e.g., leave the current ride at a suitable location and change to another connecting shuttle). In contrast, a 'classic' GUI would typically offer a step-by-step strategy (e.g., abort the current trip in a first step and then book a new ride).

As interactions with chatbots are currently often either productivity-oriented or relational, Følstad and Skjuve [18] suggest to integrate both forms to enhance conversational UX. The relational aspect, i.e., the creation of a natural and 'human-like' feeling [21] is regarded as a key challenge in designing chatbots. To achieve this, the bot's personality needs to be clearly defined [23, 47].

2.3 Evaluating UIs Along the Whole AMoD User Journey

After registration, user journeys in AMoD systems can be structured in three phases along time. For comprehensive support, the UI should assist users in all of these phases: (1) *before the ride:* the UI provides functionalities for information and booking; (2) *during ride:* the UI serves as a travel companion; and (3) *after the ride:* the UI assists with offboarding, rating the service etc. To create optimal interactions, it is necessary to be able to evaluate a UI along the whole journey. However, as existing AVs are very limited in their applicability, AMoD is still a more or less theoretical subject [39]. To counteract mentioned acceptance hurdles, it is crucial to design and evaluate UIs in this highly context-sensitive domain as early as possible. The application of context-based prototyping and evaluation methods [15, 24] are for good reason an essential ingredient in the human-centered design process.

Currently applied methods for prototyping AMoD systems can be roughly classified into three groups. At first glance, (1) real-life test tracks with actually driverless AVs (e.g., [13, 35]) seem to offer the most in terms of context-sensitivity.

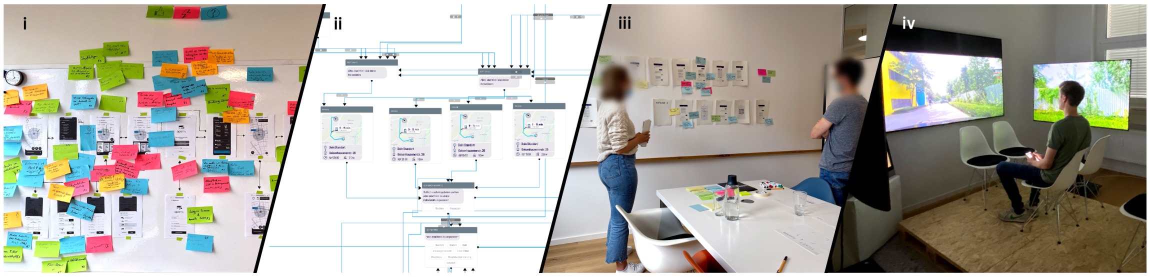

Fig. 1. Selected phases of the design process: initial conceptual designs were discussed in stakeholder workshops (i), iterated, and then used to create high-fidelity prototypes using Sketch and Botmock (ii). The GUI and chatbot prototypes were evaluated in expert studies (iii), improved, and then compared in a between-subjects user study with an immersive video-based AV simulator (iv).

However, user studies with such settings are currently only possible under very limited conditions. Test and evaluation activities are, thus, restricted to very specific scenarios (e.g., test tracks with low speed limits). (2) Wizard-of-oz settings (e.g., [27]) try to give participants the illusion of riding in an AV by hiding the human driver who is actually controlling the vehicle. A crucial aspect of such settings is to establish and to maintain the illusion of an autonomous system.

Besides 'real-world' setups like (1) and (2), lab-based evaluations using (3) driving/ride simulations provide the benefit of creating a controllable and reproducible setting with high flexibility and simplicity in data collection [10, 41]. However, because most simulator studies focus on SAE levels 0 – 3 requiring participants to control the virtual cars like they would do in reality, driving simulators are generally sophisticated constructs and only available in dedicated labs. As in AMoD passengers are not required to control the vehicle's movement, recent work [15] proposes to use rather simple video-based simulations instead of complex CGI-based setups for context-based prototyping.

3 MATERIAL AND METHOD

This work is part of a larger research project on AMoD and future mobility concepts. Within the project, an iterative human-centered design process [12] is applied. User research activities – comprising, e.g., citizen dialogues ($n = 76$), large-scale online questionnaire studies ($n_1 = 456$, $n_2 = 148$) and stakeholder interviews – are combined with literature research and (technical) requirements engineering activities to form the basis for the design of a fictitious AMoD service. Resulting artefacts like personas, user journeys, and test scenarios form the foundation for the development of various UI concepts (e.g., mobile apps and in-vehicle UIs) for interacting with a future AMoD service in a confidential and private manner.

This paper concerns the design and comparison of a GUI-based and a chatbot-based AMoD companion app. Although the two interaction concepts can be considered to be rather different, they both offer suitable approaches to interact with AMoD systems using mobile devices (section 2.2). To ensure a fair comparison of the two higher-level concepts, equivalent and representative prototypes are designed and, in a first step, evaluated in separate expert studies and then improved based on the findings. In a second step, the iterated prototypes are compared in a simulator user study.

3.1 Design Process and Prototypes

In the following, we provide details about the GUI and chatbot prototypes' design process. Both prototypes' design was optimized for an iPhone X, which was used as a test device in all three studies reported in this paper. See Fig. 1 for an overview of the design process and Fig. 2 for screenshots of the final prototype versions used in the user study.

3.1.1 GUI. At first, the GUI prototype was developed based on the mentioned preceding user research activities. After a first ideation phase, we created conceptual wireframes for an app that supports future AMoD users along the whole user journey. We used the wireframes to collect initial feedback from designers, developers, and other stakeholders (public transport providers, logistics service providers, city councils, and urban planners) in concept workshops (Fig. 1:i). Subsequently, the concept was iterated, and a high-fidelity prototype was created using Sketch v.67.

The prototype featured a map-based main view to entering travel details and requirements (e.g., departure time, destination, number of travelers, shuttle class, temporal and local flexibility). The service was designed to offer users three options for rides based on their input. The options differed, e.g., in terms of departure and travel times, shuttle classes (e.g., standard shared or express), prices, and walking times. Furthermore, the app featured a ticket wallet and an in-build navigation functionality (e.g., to find the pick-up point or the final target destination). During the ride, the GUI was designed to provide real-time travel information (e.g., current location, estimated arrival time), access to support and emergency functions, and the option to abort the current ride at the next possibility.

After optimizing the 'classic' GUI based on the findings from the expert study (Fig. 1:iii, section 4.1), it served as a foundation for the chatbot prototype.

3.1.2 Chatbot. For the design of the chatbot's personality and conversation flow, existing guidelines and recommendations [22, 23, 25, 47, 50] were considered. Besides that, an additional online-survey on CUIs for AMoD ($n = 70$) was conducted to gain insights about potential users' preferences.

Based on the survey results, the chatbot's tone of voice (queried using the four primary tone of voice dimensions of Moran [34]) was defined to be rather respectful than irreverent but balanced in terms of funny/serious, formal/casual, and enthusiastic/matter-of-fact. Regarding the bot's visual appearance, participants preferred the shuttle service provider's logo over a human, robotic or abstract avatar. Most participants preferred the chatbot to take the service's perspective when communicating (e.g., "your shuttle will be there in 5 minutes") rather than the shuttle's perspective (e.g., "I'll be there in 5 minutes"). Despite using natural language to communicate with the chatbot, participants liked the bot to provide supporting graphics and quick action buttons as shortcuts.

The resulting chatbot prototype 'AVA' (Autonomous Vehicle Assistant) was created with Botmock Conversation Designer. Like the GUI, the chatbot was also evaluated with a mixed-method expert study (Fig. 1:iii) and optimized based on the results. In a second step, we called in an immersive video-based AV simulator (Fig. 1:iv; [15]) to test and compare both concept prototypes holistically considering the complete user journey – from planning and booking over experiencing the booked ride in the simulator to reaching the target destination and rating the service quality (Fig. 3). All studies were conducted in German.

3.2 Expert Studies

Formative evaluations with HCI experts provided the basis for subsequent prototype iterations for each concept. The experts were recruited from internal design teams that were external to the project. I.e., participants did neither take part in creating the prototypes nor were familiar with the designs before participating in the study.

3.2.1 Participants. Six participants (5 female, 1 male, 0 diverse, 0 n/a) between 24 and 33 ($M = 28$) and an average working experience of five years took part in the GUI expert evaluation. The Affinity for Technology Interaction (ATI) [19] indicated rather high technology affinity among the sample with $M = 4.3$ ($SD = 0.5$; 0 = min; 6 = max).

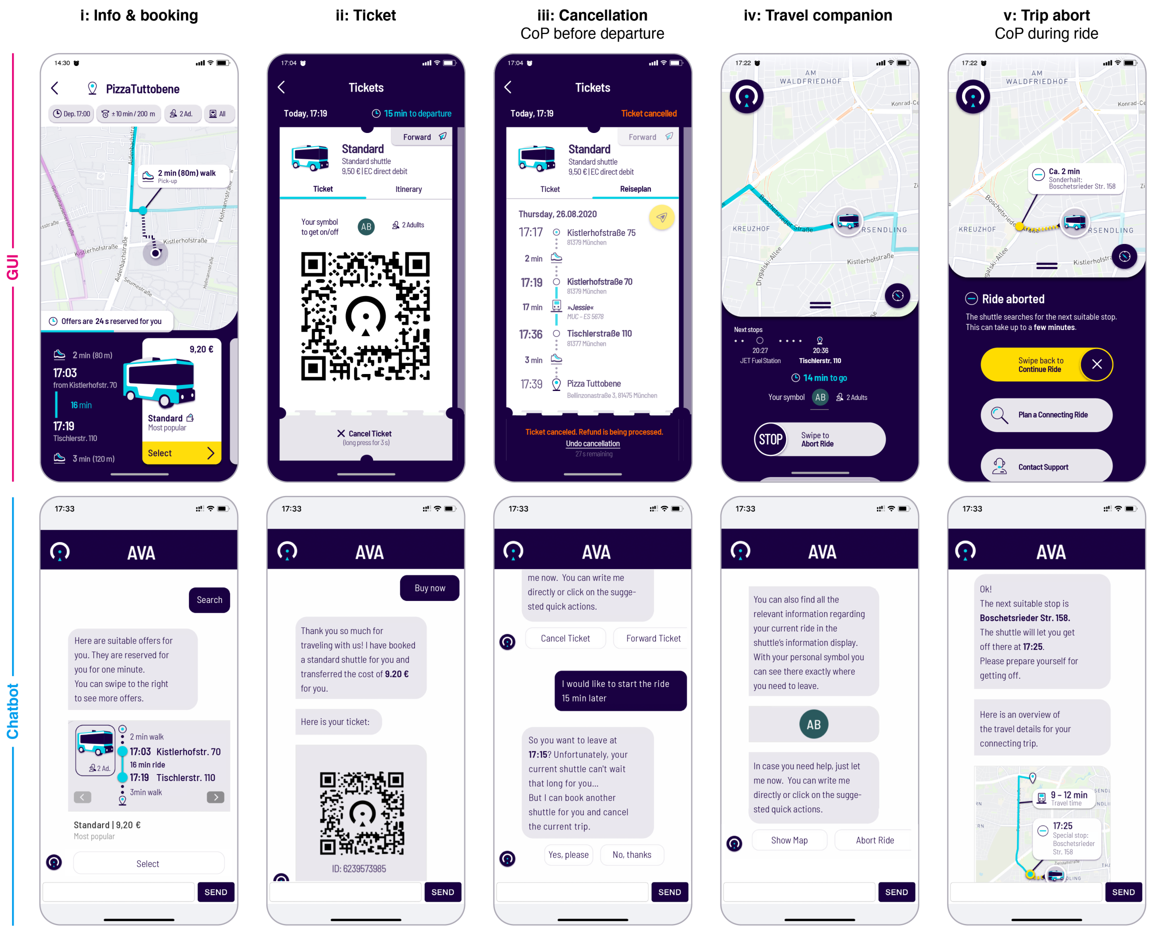

Fig. 2. Selected states of the final prototypes used in the simulator user study (translated from German to English). Top: 'classic' GUI; bottom: chatbot (text-based CUI).

In the chatbot's expert evaluation, five participants (3 female, 2 male, 0 diverse, 0 n/a) with an average age of 28 years (min: 24; max: 32) and an average working experience of about four years took part. Again, the ATI [19] indicated high technology affinity among the experts with $M = 4.7$ ($SD = 0.5$; 0 = min; 6 = max).

3.2.2 Procedure. Each expert participated in a single-session with an experimenter and a note taker. The sessions were structured in four phases: (1) pre-session interview (demographics, experience), (2) scenario- and task-based usability test with thinking-aloud and live note-taking on a whiteboard, (3) User Experience Questionnaire (UEQ; 26 bipolar items; seven-point scale; scales: Attractiveness, Stimulation, Novelty, Perspicuity, Efficiency, and Dependability) [29, 42], (4) semi-structured interview and discussion of notes (Fig. 1: iii).

Following a mixed-method approach [9], we collected both quantitative UX metrics using the UEQ and qualitative aspects by conducting a semi-structured interview and discussion at the end of the sessions to identify optimization potential. Based on the feedback and consecutive design iterations, we wanted to assure a certain quality level of the prototypes for their fair comparison in the subsequent user study.

3.3 User Study

For the comparative evaluation of the two concepts, we considered context as a crucial component to achieve valid assessments. As stated above, there are several approaches to simulate the not yet available context of using AMoD services. Since we wanted to create a reproducible experience that lets participants encounter a complete user journey in an AMoD system, we decided to go for a lab-based simulation. We adapted and enhanced the video-based method described by Flohr et al. [15] to create a straightforward mock-up of a shared AV (Fig. 1: iv; Fig. 3).

3.3.1 Experimental Design. We presented a typical use case of future AMoD systems in the user study: booking and taking a single trip in a shared AV. Here, we put special attention to two potential scenarios. (1) Journeys with 'happy paths', i.e., journeys without exceptional or error conditions and no occurring changes of plans and – since increased flexibility would be a primary reason for users to switch to AMoD [38] – (2) journeys with user-initiated changes of plans (CoP; changing the destination and departure time of an already booked trip with a shared shuttle).

Based on the insights gained from related work and our expert studies, we assumed that scenarios with 'happy paths' can be conducted without major effort or frustration using the GUI. However, we conjectured that in CoP scenarios, the strength of the chatbot's dialog-based conversation style would come into play. Here, the bot could provide a dialog-based explanation for the possible restriction (in shared rides, desired changes can only be made with consideration for the other passengers) and guide the user to a solution. Based on these assumptions, we derived the following hypotheses regarding the user acceptance of the AMoD system:

- H1.1: Acceptance is higher when using the chatbot than when using the GUI in CoP scenarios.
- H1.2: Acceptance is higher when using the GUI than when using the chatbot in 'happy path' scenarios.

As shown in other contexts [1, 7], UX is linked to user acceptance. That is why we assumed an effect analogously in UX metrics. We further assumed that different ratings in the test conditions should merely be based on task-oriented interaction quality [44] and, therefore, focused on pragmatic quality to derive the following hypotheses:

- H2.1: Pragmatic quality of the chatbot is higher than that of the GUI in CoP scenarios.
- H2.2: Pragmatic quality of the GUI is higher than that of the chatbot in 'happy path' scenarios.

To assess the hypotheses, the study used a counterbalanced mixed design with a between-subjects factor of the UI concept (GUI vs. chatbot) and a within-subjects factor of the scenario ('happy path' vs. CoP).

3.3.2 Data Collection. Quantitative and qualitative data was measured using a digital questionnaire after each ride. Acceptance was assessed using the questionnaire by Van der Laan et al. [52] (9 bipolar items with 5-stage scale; scales: Usefulness and Satisfying) as well as with the Intention to Use scale of Chen's [6] adaption of the Technology Acceptance Model (TAM) to the AV domain (2 items; five-point Likert scale). Pragmatic Quality was assessed by the respective subscales of the UEQ [29]: Perspicuity, Efficiency, and Dependability.

Another goal of the study was targeted at an exploratory evaluation. Since the UI would be the main touchpoint between the AMoD service and the users during the whole journey, we were curious how the UI concept might also affect hedonic aspects, users' emotional constitution, and their trust in the system. Therefore, we also evaluated the UEQ [29] scales Stimulation, Novelty, and Attrativeness as well as users' Trust and Emotion after each ride. The applied Trust scale (3 items; five-point Likert scale; [33]) is also part of Chen's [6] adaption of the TAM. Emotion was assessed by emotion curves drawn by participants on an adapted version of the template used by [27].

Furthermore, we intended to assess the chosen methodological approach and its suitability for evaluating AMoD systems. Therefore, we collected participants' presence perception using the Igroup Presence Questionnaire (IPQ; 14 items;

seven-point Likert scale; [45, 46]) and a single item (five-point Likert scale; *'I found the ride in the simulator realistic.'* [15]) as well as participant's Wellbeing (single item with five-point Likert scale; *'I felt comfortable during the ride.'* [15]), and the occurence of Simulator Sickness (single item with five-point Likert scale: *'I felt nauseous during the ride in the simulator.'* [20]). A semi-structured interview on the AMoD service, the UI concepts, and the simulation closed the evaluation session.

3.3.3 Participants. To achieve sufficient power ≥ .80 with an alpha error rate of $\alpha \leq .05$ for calculating inferential statistics (planned contrasts and ANOVAs with repeated measures and within-between interactions), a required sample size of 34 participants was determined a-priori using G*Power 3.1 for Mac. Considering practical impacts, medium effects according to Cohen [8] were assumed.

In total, 35 participants took part in the study. However, one participant's data had to be excluded from the analysis due to prototype and simulation errors during the session. This resulted in a sample of $n = 34$ participants (15 female, 18 male, 0 diverse, 1 n/a) between the ages of 19 and 61 ($M = 29.9$, $SD = 9.8$). The ATI short scale (ATI-S) [54] indicated rather high technology affinity among the sample with $M = 4.2$ ($SD = 1.3$; 0 = min; 6 = max). All participants were recruited via online postings and received financial compensation (30 €).

3.3.4 Procedure. Prior to the study, participants were randomly (counterbalanced) assigned to one of the between-subjects groups leading to 17 participants testing the chatbot and 17 participants testing the GUI.

At the beginning of the study session, participants received a detailed briefing on the study's purpose and procedure and the potential side effects of simulator sickness. Afterwards, they signed an informative participation consent form and filled out a pre-questionnaire (demographics, ATI-S). Participants were encouraged to think aloud and ask questions whenever they wanted and – in case they felt at unease – to pause or quit the study at any time without consequences.

For the task-based testing, participants experienced two AMoD journeys in the lab-based setup. The order of the journeys was randomized (counterbalanced). For each of the journeys, they were provided with a contextual scenario. Their task was to travel with a friend to a public park (journey 1) and back home (journey 2) using the AMoD service. Therefore, they had to enter the respective target location and departure time and select one out of the three options provided by the app.

After booking, participants used the app's build-in navigation functionality, checked in at a paper-prototyped door UI using the received ticket QR code, and took the ride in the AV simulator. In the 'happy path' journey, participants did not receive explicit instructions on what to do during the ride. I.e., they were free to, e.g., monitor the information display or to just enjoy the (simulated) ride. In the CoP condition, two CoPs occurred. (1) A change of the departure time for an already booked ride implied canceling the initial booking prior to departure and booking a new ride. (2) A change of the target destination during the ride required participants to abort the ongoing ride and book a connecting ride to the new destination (a restaurant). When they reached the target destination ('happy path'; after a 14 min ride) respectively the special stop (CoP; after a 7 min ride), participants checked-out and rated the service using the app (Fig. 3).

After each ride, participants assessed acceptance, UX, and trust using a digital questionnaire and reported their emotional changes during the journey by drawing an emotion curve. At the end of the session, participants evaluated presence perception, wellbeing, and simulator sickness with the digital questionnaire and took part in a semi-structured interview. Each session took about 75 to 90 minutes and was conducted by one experimenter and one note-taker.

3.3.5 Apparatus. For a context-based evaluation of the concepts, we adapted the immersive video-based driving simulator proposed by [15]. In contrast to the setup of [15], large TV screens with 4K resolution instead of projectors formed the simulator's basis. The video footage was recorded with three action cameras (left, center, right) while driving through an urban environment and supplemented with audio footage (e.g., opening/closing doors, signal sounds, voice prompts). To

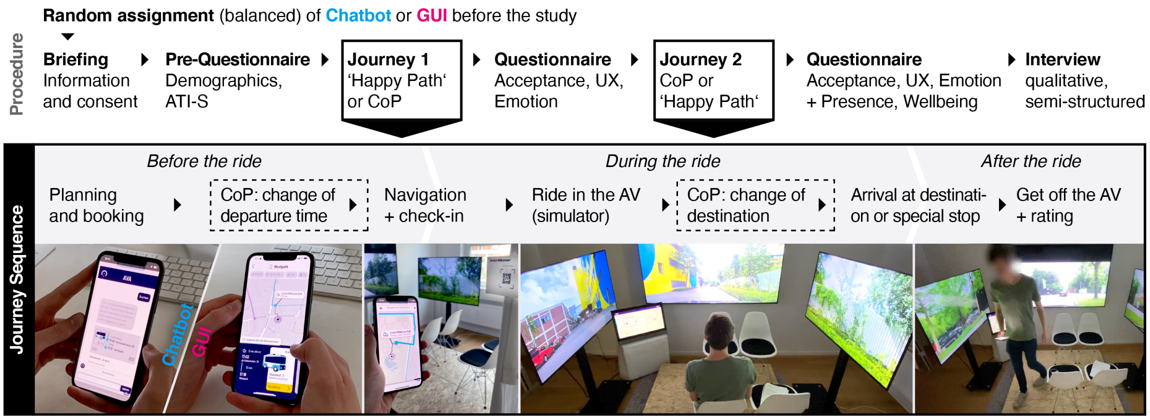

Fig. 3. Study procedure and journey sequence illustrating how participants experienced the prototypes in the immersive video-based lab setup. Scenario order ('Happy Path' and then CoP or vice versa) has been randomly assigned (counterbalanced).

enhance the simulation's immersion, the screens were configured to form a closed space embracing an elevated platform with a seating group (Fig. 1: iv; Fig. 3). A 26-inch in-vehicle information display – illustrating a timetable, upcoming stops, and a map-view with the AV's current position and planned route – complemented the setup and the UI concepts.

4 RESULTS

Created prototypes were evaluated in mixed-method expert studies (n_{GUI} = 6; $n_{Chatbot}$ = 5) and improved based on the findings (i.e., discovered usability problems were solved and the experts' suggestions for improvement were implemented). In a second step, the iterated prototypes were compared in a between-subjects simulator study (n = 34).

4.1 Expert Studies

Both the GUI and the chatbot prototypes scored above average to excellent ratings in the UEQ subscales [29, 42] (scales range from -3 to 3). Qualitative findings revealed minor and cosmetic usability problems.

4.1.1 GUI. According to the UEQ Benchmark [43], the GUI received excellent ratings in terms of Attractiveness (*M (SD)* = 2.1 (0.7)), Efficiency (*M (SD)* = 2.0 (0.7)), and Dependability (*M (SD)* = 1.8 (0.4)) as well as good ratings for Perspicuity (*M (SD)* = 1.9 (0.3)), Stimulation (*M (SD)* = 1.5 (0.7)), and Novelty (*M (SD)* = 1.4 (1.1)) by the six experts.

The use of commonly established visual and conceptual patterns (e.g., form and navigation design, or auto-suggestions while entering start and target locations) was commended by three experts. Several non-standard functions (e.g., reservation time, trip abort) were, however, described by five experts to be ambiguous since the GUI lacked explanations.

One participant questioned whether specific functions (e.g., search for new rides) need to be always visible (e.g., during the ride) and recommended an adaptive behavior with respect to the context of use. Four experts considered the quick access to emergency functionalities during the ride to be useful, as it, e.g., supports a feeling of safety.

The map-based functionalities (e.g., navigation, context-based information) were positively mentioned by five experts. P1 and P4 especially considered the visualization of various ride options on the map to be helpful for understanding the differences between the offers. However, some experts (P2, P3) found the search results to be crowded and preferred fewer details. P3 would have liked to have all detailed information accessible on-demand, e.g., by expanding the journey overview.

While walking to the pick-up point, i.e., shortly before the departure time of the AV, P5 and P6 would want the UI to communicate how long the AV would wait for the passenger at the pick-up. P3 wanted the GUI to remember changes in filter settings for the next booking. In change of plans situations, P2 and P5 wanted the UI to be capable of making adjustments to already booked trips. However, in shared rides, this is conceptually restricted, as changes would also affect other passengers. When aborting a current trip, P2 would have expected that map-based functionalities are still accessible and that information on onward travel options get provided in order to proactively support the user. Overall, the GUI's visual design was described as appealing by five of the six experts. Further optimization potential was found regarding the design of single components and details (e.g., icon metaphors, button design).

4.1.2 Chatbot. The chatbot received excellent ratings for Stimulation (*M (SD)* = 1.8 (0.9)), and Novelty (*M (SD)* = 1.7 (0.9)) as well as above average ratings for Attractiveness (*M (SD)* = 1.6 (0.9)), Perspicuity (*M (SD)* = 1.7 (0.8)), and Efficiency (*M (SD)* = 1.4 (1.0)), but a below average rating for Dependability (*M (SD)* = 1.1 (1.1)) [43]. The latter indicates lack of predictability and feeling of control.

All five experts generally liked the conversational approach for interacting with the autonomous system. Three saw major advantages in the assistive and guiding nature of the chatbots' conversation flow and the high efficiency when users become experienced with the interaction. P2 especially appreciated the option to set all parameters at a time with a single message. However, this advantage was assessed to be less pronounced when users are not familiar with the chatbot. To counteract this issue, the experts suggested to optimize the on-boarding process.

All experts considered the chatbot's tone of voice to be appropriate to the bot's capabilities and role. P3 and P5 explicitly liked the neutral and friendly conversation style. However, two experts suggested to make the language even more human and to improve system feedback by using more relational messages.

P2, P3, and P4 recommended to combine chatbots with classic GUIs, i.e., to create a 'hybrid' UI with aspects of both. This idea was indirectly also emphasized by other experts since all participants would have wanted to always have access to common functions like, e.g., the map or tickets. P2, P4, and P5 would have wanted to directly manipulate set parameters in the messages sent by the bot as it is possible in classic GUIs. P1, P2, and P3 argued that the visual design of quick actions and buttons was not salient enough. Further optimization potential also concerned ambiguous or unclear wordings.

4.2 User Study

Descriptive and inferential statistics were calculated using JASP 0.12, jamovi 1.2, and R Studio 1.2 (R 4.0). Planned contrasts were performed to answer the hypotheses in terms of a potential underlying interaction effect. Repeated measurement ANOVAs were calculated to explore significant differences related to the used UI concept. Qualitative data from the interviews, session notes, and anecdotal evidence was digitally collected, structured, and analyzed.

4.2.1 Acceptance. Usefulness, Satisfying, Intention to Use, and Trust scales score high ratings in all conditions (Fig. 4). While planned contrasts did not reveal the expected interaction effect in terms of Usefulness ($t(32) = -0.08$, $p = .398$; $d = -0.42$) and Satisfaction ($t(32) = -0.17$, $p = .863$; $d = 0.08$), a significant large interaction effect for Intention to Use ($t(32) = 2.99$, $p = .005$; $d = 1.45$) was revealed. In other words, participants reported a higher intent to use the GUI in 'happy path' scenarios, but preferred the chatbot in scenarios with CoPs.

However, in general, the service was rated significantly more satisfying when using the GUI than when using the chatbot ($F(1,32) = 5.25$, $p = .029$; $\eta^2_G = 0.11$). In terms of Usefulness and Intention to Use no such effect was found. The AMoD service also received high ratings for both UI concepts in terms of Trust without a meaningful difference between the scenario conditions ($F(1,32) = 1.90$, $p = .178$; $\eta^2_G = 0.053$).

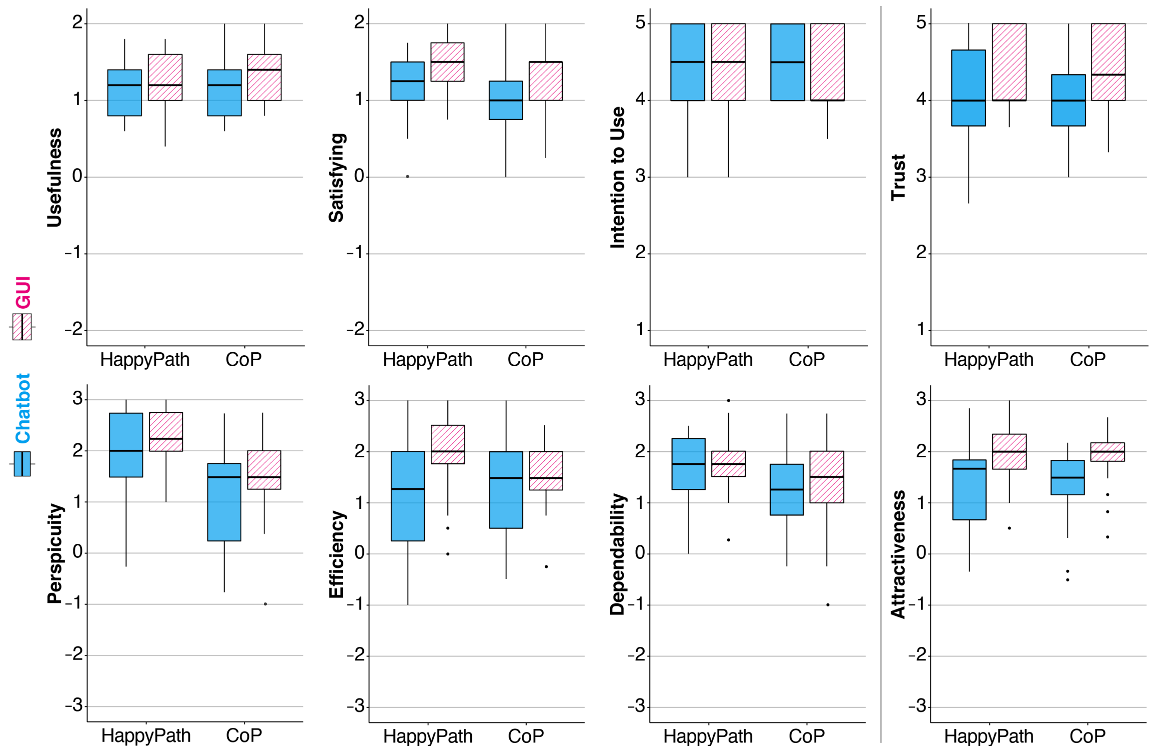

Fig. 4. Boxplots of acceptance and UX scales from the user study; $n_{Chatbot}$ = 17; n_{GUI} = 17.

4.2.2 UX. Regarding the pragmatic quality, both concepts received high ratings in terms of Perspicuity, Efficiency, and Dependability (Fig. 4). Planned contrasts did not reveal significant interaction effects – neither for Perspicuity ($t(32) = -0.04$, $p = .966$; $d = -0.02$), Efficiency ($t(32) = -1.23$, $p = .227$; $d = -0.60$), nor for Dependability ($t(32) = 0.26$, $p = .796$; $d = 0.13$). While no meaningful differences in terms of the UI concept could have been found in the pragmatic quality scales, an ANOVA revealed significantly higher ratings for the GUI than for the chatbot and a medium effect in terms of Attractiveness (*M(SD)*: $\text{GUI}_{\text{HappyPath}}$ = 2.0 (0.6), $\text{Chatbot}_{\text{HappyPath}}$ = 1.4 (0.9), GUI_{CoP} = 1.8 (0.6), $\text{Chatbot}_{\text{CoP}}$ = 1.3 (0.8); $F(1,32) = 5.56$, $p = .025$; $\eta^2_G = 0.13$). In other words, participants rated the overall impression of the service better when a GUI was used. Regarding the hedonic qualities Stimulation and Novelty no significant differences were found (*M(SD)* for Stimulation: $\text{GUI}_{\text{HappyPath}}$ = 1.0 (0.9), $\text{Chatbot}_{\text{HappyPath}}$ = 0.9 (0.7), GUI_{CoP} = 1.5 (0.7), $\text{Chatbot}_{\text{CoP}}$ = 1.1 (0.6); *M(SD)* for Novelty: $\text{GUI}_{\text{HappyPath}}$ = 1.1 (1.0), $\text{Chatbot}_{\text{HappyPath}}$ = 0.7 (0.9), GUI_{CoP} = 1.1 (0.8), $\text{Chatbot}_{\text{CoP}}$ = 1.0 (0.5)).

4.2.3 Emotion. Emotional curves indicate quite positive feelings throughout the journey and across conditions. For a comparative evaluation, the curves were digitized, normalized to 'before use', and stacked (opacity: 0.1) for each condition (Fig. 5). The stacked curves reveal an overall positive emotion trajectory throughout the 'happy path' journey with the GUI. In the 'happy path' with the chatbot, they show larger fluctuations and some negative peaks during the ride and in the booking phase. During the booking phase and before the ride, almost all curves appear to follow a positive trend. For both the GUI and the chatbot, the trend seems to be slightly more positive in the CoP than in the

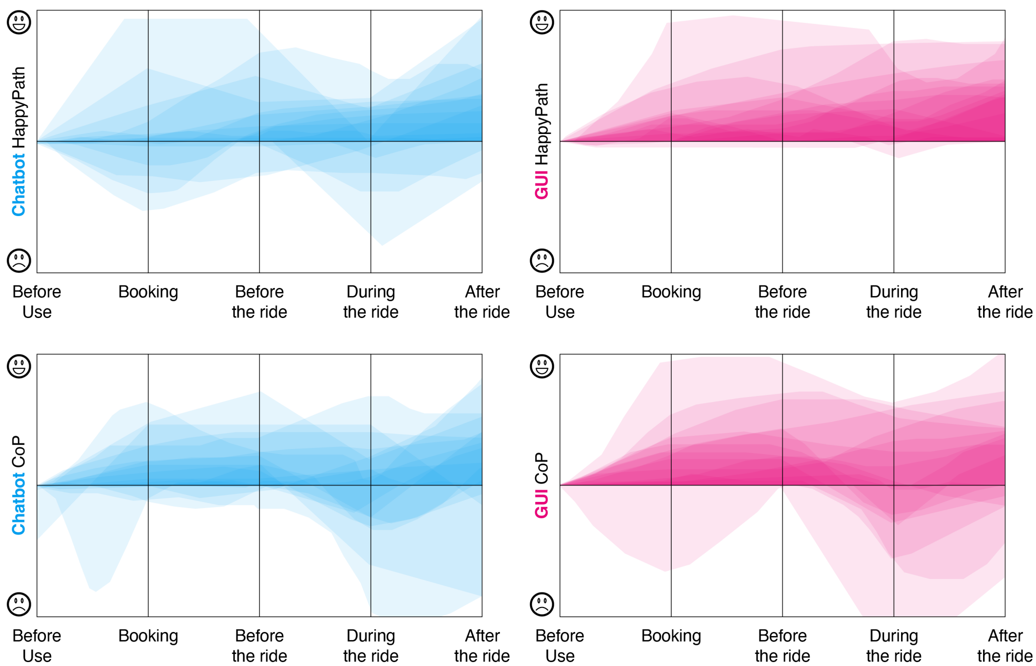

Fig. 5. Stacked emotion curves (opacity: 0.1) drawn by the participants after each journey, normalized at 'before use'. Top left: chatbot 'happy path' (n = 17), top right: GUI 'happy path' (n = 17), bottom left: chatbot CoP (n = 17), bottom right: GUI CoP (n = 17).

'happy path' conditions. I.e., changing the departure time seems to be a positive experience with both concepts. During the ride, participants indicate adverse effects in terms of their emotional constitution in all conditions except for the GUI's 'happy path'.

4.2.4 Presence, Wellbeing, and Simulator Sickness. IPQ scales scored medium to high ratings (Fig. 6). While Involvement and Experienced Realism fell rather short with medium ratings, Spatial Presence and the General scale achieved above middle ratings (Fig. 6). Four participants explicitly described the simulated ride as "realistic". P4, P17, and P28 commend the driving style of the simulated AV (e.g., "I feel relaxed", "I feel safe"). P2, P4, and P16 found it unrealistic that there was always a free parking slot available. Thirty of the 34 participants left the AV (i.e., the simulator) on their own when the simulation reached the target destination, while four remained seated. Similar to the IPQ's general scale, the single item *"I found the ride in the simulator realistic."* (Fig. 6) achieved a high rating with a median of 4. Participants' did not seem to encounter adverse effects during the simulation, which is illustrated by a positive assessment of Wellbeing (median = 4) and low occurrence of Simulator Sickness (median = 1; see Fig. 6).

4.2.5 Qualitative Findings: UI Design. Six GUI participants wanted to have a "direct" option to make adjustments to booked trips, e.g., change departure time without cancellation. P32 wanted the GUI to be more supportive ("take me by the hand") in CoP situations. Two participants preferred to have an option for speech-based in- and output. In general, some participants applied a 'trial and error" strategy on the GUI prototype, i.e., when they did not know what to do, they just tapped randomly.

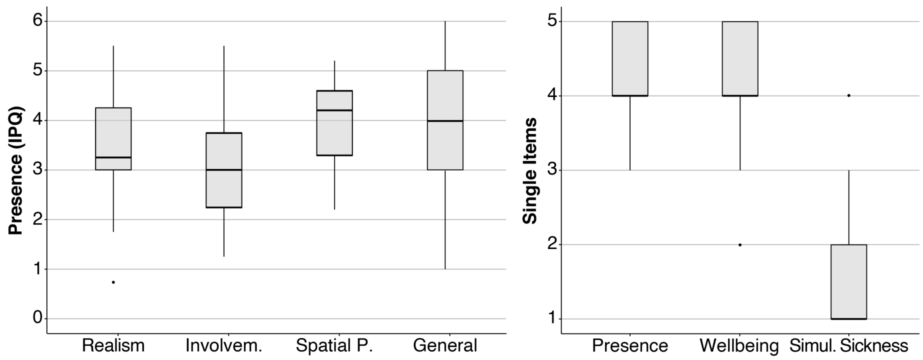

Fig. 6. Left: boxplots of the four IPQ scales Realism, Involvement, Spatial Presence and General (0 = min, 6 = max); Right: boxplots of the assessed single items Presence, Wellbeing, and Simulator Sickness (1 = min, 5 = max); n = 34.

While all participants testing the GUI considered the type of the communication to be appropriate, three participants testing the chatbot either generally did not like bots (P4, P24) or found it inconvenient to articulate their needs using written text (P10). Several users (P12, P13, P23) feared that the chatbot would not understand them correctly if they would not use specific keywords. In contrast, P9 commended that the chatbot understands whatever they said. The qualitative assessment of the chatbot's tone of language confirmed the intended balance in terms of enthusiastic/matter-of-fact, formal/casual and funny/serious, as well as a rather respectful than irreverent way (see [34]).

Besides that, some participants would have wanted the chatbot to call them by their name (P10) and to make smalltalk (P12, P24). Five participants wanted the messages of the chatbot to be shorter – especially when they would use it more often. P11, P24, and P32 commended the chatbot's supporting nature, e.g., departure reminder. Five participants suggested to add direct touch interactions to the chatbot, e.g., to change values by clicking on chatbot messages. Further suggestions concerned speech output (P12), an always available "context view" (P7), and a app menu as used in classic GUIs (P4). Similar to the latter, P11 preferred a combination of both GUI and chatbot.

4.2.6 Qualitative Findings: AMoD Service Design and Simulation. Two participants (P25, P32) suggested to introduce gamification elements, e.g., a calculation of saved CO_2 using a shared AV compared to taking the own car. During the ride in the simulator 15 participants (7 using the GUI, 8 using the chatbot) stated that they did not have sufficient information. Seven of them (4 using GUI, 3 using chatbot) referred to a specific CoP situation where the in-vehicle display did not provide required information. This illustrates the crucial connection of various service touchpoints (e.g., mobile app, in-vehicle display) and the context.

Two participants (P13, P30) noted that they were bored during the simulated ride in the AV. P2 would have wanted to use his smartphones as they would typically do while riding in public transport or taxis. Several participants wondered how the AV would behave if, e.g., a passenger would not leave at the designated destination (P7, P17, P20), a passenger without a valid ticket entered the AV (P28), or an emergency occurred (P6, P28). P6 found it "embarrassing" if all other passengers in a shared AV would know who requested an extra stop.

5 DISCUSSION

The conducted expert and user studies reveal the advantages and disadvantages of both chatbots and 'classic' GUIs – in general, and specifically for certain AMoD scenarios. Table 1 situates and contrasts these among the findings from previous work. In the following sections, we discuss how they relate to our research question regarding the UI concept's effect on acceptance and UX in 'happy path' and CoP scenarios. We also discuss the applied context-based prototyping approach [15], limitations, and future work.

Table 1. Summary of general advantages (+) and disadvantages (–) of 'classic' GUIs and chatbots based on the findings of related work and our expert (E) and user (U) studies.

	'Classic' GUIs	Chatbots
+	High discoverability through visual clues [32], (U)	Intuitive input via natural language [30], (E), (U)
	Support users' mental model [32, 49], (E)	Expression of needs with own words [49, 50], (E), (U)
	Easy understanding of system scope [32, 49], (U)	Efficient for experienced users [49], (E)
	Suitable for displaying loads of information [5, 49]	Adaptability and personalization [17, 56], (E)
	Direct manipulation of elements (E), (S)	Ease of use [51], assistance and guidance (E), (U)
	Efficient shortcuts and established concepts [50], (E), (U)	Direct clarification of ambiguities [49]
	Potentially (more) attractive visual appearance (E)	Simplifiable information search process [51]
–	Users need to understand the design [49, 50], (E), (U)	Interpretation problems and misunderstandings [17], (U)
	Ambiguities of non-standard functions (E), (U)	Mentally demanding articulation of needs [32, 49], (U)
	Restrictive in terms of alternative usage ways (U)	Step-wise learning of system capabilities [49], (E)
	Potential information overload and distraction (E)	User requests can exceed system scope [50], (U)

5.1 GUI vs. Chatbot for AMoD Systems

Generally, both 'classic' GUIs and chatbots hold several advantages and disadvantages. As illustrated in Table 1, our results confirm related works' findings on GUIs and chatbots and complement the existing literature body with some new insights and a comparative case study on the two concepts for the AMoD domain. Across our studies, both the 'classic' GUI and the chatbot received high acceptance and UX ratings. Thus, the results confirm that both concepts can be regarded as valid and suitable options to consider when designing mobile applications for AMoD systems.

In the user study, the GUI was reported to be significantly more satisfying and attractive than the chatbot in general (both with medium-sized effects [8]). Descriptive plots of other acceptance and UX scales (Fig. 4) support this tendency but do not reveal meaningful differences. The very positive assessment of the GUI's visual design in the expert and the user study might, however, be co-responsible for its higher Attractiveness ratings. While the chatbot prototype still achieved a relatively high assessment, it probably fell shorter in terms of visual appearance compared to the GUI because of restricted design possibilities. The general pros and cons of the concepts (Table 1) hold further possible explanations for the effects.

Planned contrasts revealed a significantly higher Intention to Use the GUI in 'happy path' scenarios but the chatbot in CoP scenarios. This is partly supported by the GUI's emotion curves (Fig. 5). They revealed positive experiences throughout the 'happy path' scenarios with the GUI but relatively large derivations in CoP scenarios. However, the chatbot's emotion curves (Fig. 5) and other acceptance scales (Usefulness, Satisfying, Trust; Fig. 4) do not back this observation. Consequently, we have to reject the acceptance-related hypotheses H1.1 and H1.2. Similarly, we have to reject H2.1 and H2.2 as none of the UEQ's pragmatic quality scales returned the expected interaction pattern.

In both chatbot studies, participants preferred to directly manipulate parameters (e.g., date, time, destination) displayed by the bot. While this can be regarded as a clear benefit of GUIs in general (Table 1), it should be considered whether it might also be a usable option to implement direct manipulability in chatbots – e.g., form elements, date pickers, app menus. Alternatively, since several non-standard functions and information in the GUI have not been 'instantly' understood in both the expert study and the simulator user study, we recommend supplementing the 'classic' GUI with conversational elements. The assistive nature of chatbots (and CUIs in general), their linear conversation flow, and the on-demand provision of adequate explanations seem to be an excellent addendum to overcome the associated challenges of new and non-standard functions for future AMoD systems.

5.2 Context-based Prototyping with Immersive Video

As AMoD is still a theoretical matter [39], we considered simulating holistic user journeys as a crucial component for successful evaluation and, thus, for developing adequate human-AV interactions that can counteract acceptance hurdles.

The video-based approach adapted from [15] generally provided a straightforward and controllable environment that made it possible to immerse participants in the context in order to let them experience a complete AMoD journey. Assessment of the IPQ scales Realism, Involvement, Spatial Presence, and its General scale supports this with medium to high scores, along with qualitative findings. Although our TV-based setup differed from the projection-based setup of [15], the results are quite comparable. This indicates a successful replication of the approach. The very positive assessment of the single items on presence and wellbeing as well as the low occurrence of simulator sickness symptoms further supports the general suitability of the method.

Since people seemed to get bored doing nothing during the (simulated) ride – which might have affected participant's assessment, e.g. in terms of emotion curves – it might be beneficial for further empirical studies to introduce secondary tasks (e.g., attention tests, monitoring of changes in the UI, or simply reading a book).

5.3 Limitations

Some limitations of this work should be noted. In general, this paper focuses on mobile applications for AMoD. Consequently, the generalisability of the reported studies and the comparative evaluation of GUIs and chatbots is limited.

Created prototypes were optimized for following a pre-defined scenario. However, both prototypes featured alternative interaction paths to some extend. E.g., in the CoP scenario 'change of destination', the chatbot offered a step-by-step approach similar to the GUI but was also able to provide a context-based shortcut if the participant directly requested to change the destination via text. While such alternative usage ways incorporated the concepts' benefits into the study, they might also affect the study's reliability.

The GUI was created using Sketch's prototyping features which highlighted interactive elements on tap. As a result, participants applying a 'trial and error' procedure received hints from the prototype on how to proceed. The chatbot prototype made use of the language understanding capabilities provided by Botmock. Though we added an extensive collection of likely intents and formulations, not all possible formulations were anticipated. Thus, some participants encountered 'dead ends' during the study and had to try different formulations. In some cases, this resulted in a trial and error behavior as well. Consequently, both prototypes' levels of fidelity might have impacted the results.

Participants in all three studies reported a rather high affinity for technology interaction. While this is considered to be a common phenomenon in HCI studies [19], it might impair external validity. Furthermore, having a look at the very positive evaluation of both prototypes and scale constitutions, ceiling effects cannot be excluded.

The user study was conducted in the late summer of 2020, i.e., during the COVID-19 pandemic. Following regulations and recommendations of local and national authorities, we applied several precautions, such as distancing and hygiene measures. While we consider the effects on the study's actual conduct to be minor, the situation might have affected the sample composition, as, e.g., only people with low fear of the situation might have signed up for the study.

5.4 Future Work

Similar to the recommendation of [18] to integrate relational and productivity-oriented interaction for conversational design, future work should consider transferring this approach in terms of combining GUI-based with CUI-based concepts to create flexible and accessible interactions. Such could be text-based but, considering accessibility requirements, also speech- or gesture-based. Like 'hybrid' UIs in today's cars, the approach could combine the benefits of both worlds.

Future work should investigate appropriate use cases and create design guidelines for such 'hybrid' AMoD UIs. In doing so, different in- and output modalities should be considered – also making use of further technological advances like, e.g., text-to-speech conversion and vice versa. In terms of the system's accessibility, this could be used to integrate specific demographic groups' needs – e.g., blind people or older adults might feel more confident talking to an AMoD system instead of chatting or tapping.

Furthermore, as some participants suggested, integrating gamification elements (e.g., calculating saved CO_2) might hold attractive potentials for future designs to increase the system's acceptance and hedonic quality.

6 CONCLUSION

Both GUIs and chatbots come with their specific pros and cons. Based on our findings, we derive three design recommendations for human-AMoD interaction. (1) Use 'classic' GUIs as a basis for mobile interaction. In the direct comparison, the GUI was reported to be generally more attractive and satisfying. GUIs also seem to be superior in standard use cases with 'happy path' scenarios. However, the conversation with a chatbot-based agent can increase users' intention to use the system and support users in demanding scenarios. Therefore, we recommend to (2) integrate (text-based) conversational elements in GUI-based mobile applications where appropriate to enhance user control and facilitate error recovery. As AMoD is still a theoretical subject [39], we further recommend to (3) use context-based prototyping [15, 24] from early development phases on to consider environmental factors and the interdependence of various touchpoints (e.g., mobile apps and in-vehicle displays). Video-based simulation [15] provides a suitable and straightforward basis to consider all stages of the AMoD user journey holistically. Future work should investigate applying these recommendations to other usage scenarios and the creation of design guidelines for accessible AMoD systems.

ACKNOWLEDGMENTS

This research has 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. The authors thank the study participants as well as their colleagues and project partners for their support. Furthermore, the authors thank the anonymous reviewers for their time and helpful feedback.

REFERENCES

[1] Tony Ahn, Seewon Ryu, and Ingoo Han. 2007. The impact of Web quality and playfulness on user acceptance of online retailing. *Information and Management* 44, 3 (apr 2007), 263–275. https://doi.org/10.1016/j.im.2006.12.008

[2] Wolfgang Bernhart and Stefan Riederle. 2019. Mobility's great leap forward. A world on the verge of autonomy. *Roland Berger ´s Automotive Disruption Radar* 5, March (2019), 14.

[3] Hannah Biermann, Ralf Philipsen, Teresa Brell, and Martina Ziefle. 2020. Shut Up and Drive? User Requirements for Communication Services in Autonomous Driving. In *Proceedings of the 22nd International Conference on Human-Computer Interaction (HCII '20)*. Springer International Publishing, 3–14. https://doi.org/10.1007/978-3-030-50523-3_1

[4] Teresa Brell, Hannah Biermann, Ralf Philipsen, and Martina Ziefle. 2019. Conditional privacy: Users' perception of data privacy in autonomous driving. In *Proceedings of the 5th International Conference on Vehicle Technology and Intelligent Transport Systems (VEHITS '19)*. 352–359. https://doi.org/10.5220/0007693803520359

[5] Raluca Budiu. 2018. The User Experience of Chatbots. https://www.nngroup.com/articles/chatbots/

[6] 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

[7] Vincent Cho, T. C.Edwin Cheng, and W. M.Jennifer Lai. 2009. The role of perceived user-interface design in continued usage intention of self-paced e-learning tools. *Computers and Education* 53, 2 (sep 2009), 216–227. https://doi.org/10.1016/j.compedu.2009.01.014

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

[9] John W. Creswell. 2014. *A Concise Introduction to Mixed Methods Research.* Sage Publications, Thousand Oaks, California, USA. https://books.google.nl/books?hl=nl{&}lr={&}id=51UXBAAAQBAJ{&}oi=fnd{&}pg=PR1{&}dq=creswell+mixed+methods+notation{&}ots=69GpJ4RsNA{&}sig=RS6mWCewEIvBXq85S48z1zOPPVM

[10] 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 2012)*, Andrew Spink, Fabrizio Grieco, Olga Krips, Leanne Loijens, Lucas Noldus, and Patrick Zimmerman (Eds.). Noldus Information Technology, Utrecht„ 47–50.

[11] DIN. 2008. Ergonomics of human-system interaction – Part 171: Guidance on software accessibility (ISO 9241-171:2008) English version of DIN EN ISO 9241-171:2008-10. , 97 pages. https://doi.org/10.31030/1426650

[12] DIN. 2011. Ergonomics of human-system interaction – Part 210: Human-centred design for interactive systems (ISO 9241-210:2010): English translation of DIN EN ISO 9241-210:2011-01. https://doi.org/10.31030/1728173

[13] Verena Distler, Carine Lallemand, and Thierry Bellet. 2018. Acceptability and acceptance of autonomous mobility on demand: The impact of an immersive experience. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems (CHI '18)*, Vol. 2018-April. 1–10. https://doi.org/10.1145/3173574.3174186

[14] Grace Eden. 2018. Transforming cars into computers: Interdisciplinary opportunities for HCI. In *Proceedings of the 32nd International BCS Human Computer Interaction Conference (HCI '18)*. https://doi.org/10.14236/ewic/HCI2018.73

[15] 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 Designing Interactive Systems Conference (DIS '20)*. Association for Computing Machinery. https://doi.org/10.1145/3357236.3395468

[16] Asbjørn Følstad and Petter Bae Brandtzæg. 2017. Chatbots and the New World of HCI. *Interactions* 24, 4 (2017), 38–42. https://doi.org/10.1145/3085558

[17] Asbjørn Følstad and Petter Bae Brandtzaeg. 2020. Users' experiences with chatbots: findings from a questionnaire study. *Quality and User Experience* 5, 3 (2020), 1–14. https://doi.org/10.1007/s41233-020-00033-2

[18] Asbjørn Følstad and Marita Skjuve. 2018. Business and pleasure? Relational interaction in conversational UX. In *CHI' 18 workshop: Voice-based Conversational UX Studies and Design*. https://sintef.brage.unit.no/sintef-xmlui/handle/11250/2500859

[19] 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

[20] Peter J. Gianaros, Eric R. Muth, J. Toby Mordkoff, Max E. Levine, and Robert M. Stern. 2001. A Questionnaire for the Assessment of the Multiple Dimensions of Motion Sickness. *Aviat Space Environ Med.* 72, 2 (2001), 115–119.

[21] Ulrich Gnewuch, Stefan Morana, Marc T.P. Adam, and Alexander Maedche. 2018. Faster is not always better: Understanding the effect of dynamic response delays in human-chatbot interaction. In *Proceedings of the 26th European Conference on Information Systems: Beyond Digitization - Facets of Socio-Technical Change, ECIS 2018*. 17.

[22] Google. 2020. Conversation design | Dialogflow and legacy Actions SDK. https://developers.google.com/assistant/conversational/df-asdk/design

[23] Erika Hall. 2018. *Conversational Design.* A Book Apart, Sebastopol, CA. https://learning.oreilly.com/library/view/conversational-design/9781492017950/

[24] Marius Hoggenmueller, 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 (CHI '21)*. https://doi.org/10.1145/3411764.3445159

[25] IBM. 2020. IBM | Conversational UX: Introduction. http://conversational-ux.mybluemix.net/design/conversational-ux/

[26] 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

[27] 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

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

[29] Bettina Laugwitz, Theo Held, and Martin Schrepp. 2008. Construction and evaluation of a user experience questionnaire. In *Proceedings of the 4th Symposium of the Workgroup Human-Computer Interaction and Usability Engineering of the Austrian Computer Society (USAB'08)*, Vol. 5298. 63–76. https://doi.org/10.1007/978-3-540-89350-9_6

[30] Michael McTear, Zoraida Callejas, and David Griol. 2016. *The Conversational Interface: Talking to Smart Devices*. Springer International Publishing, Basel, Switzerland. 422 pages. https://doi.org/10.1007/978-3-319-32967-3

[31] Michael F. McTear. 2017. The Rise of the Conversational Interface: A New Kid on the Block? In *Future and Emerging Trends in Language Technology – Machine Learning and Big Data*, José F. Quesada, Francisco J. Martín Mateos, and Teresa López-soto (Eds.). Springer, Cham, 38–49. https://doi.org/10.1007/978-3-319-69365-1

[32] Alex Mohebbi. 2018. Why Conversational Interfaces are taking us back to the Dark Ages of Usability. https://uxdesign.cc/why-conversational-interfaces-are-taking-us-back-to-the-dark-ages-of-usability-fa45fefb446b

[33] Mareike Möhlmann. 2015. Collaborative consumption: determinants of satisfaction and the likelihood of using a sharing economy option again. *Journal of Consumer Behaviour* 14 (2015), 193–207. https://doi.org/10.1002/cb.1512

[34] Kate Moran. 2016. The Four Dimensions of Tone of Voice. https://www.nngroup.com/articles/tone-of-voice-dimensions/

[35] 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

[36] Christina Pakusch, Johanna Meurer, Peter Tolmie, and Gunnar Stevens. 2020. Traditional taxis vs automated taxis – Does the driver matter for Millennials? *Travel Behaviour and Society* 21 (oct 2020), 214–225. https://doi.org/10.1016/j.tbs.2020.06.009

[37] Marco Pavone. 2016. Autonomous Mobility-on-Demand Systems for Future Urban Mobility. In *Autonomous Driving – Technical, Legal and Social Aspects*, Markus Mauerer, J. Christian Gerdes, Barbara Lenz, and Hermann Winner (Eds.). Springer Nature, Berlin, Heidelberg, Chapter 19, 387–404. https://doi.org/10.1007/978-3-662-48847-8

[38] Ralf Philipsen, Teresa Brell, Hannah Biermann, and Martina Ziefle. 2020. On the Road Again - Explanatory Factors for the Users' Willingness to Replace Private Cars by Autonomous on-Demand Shuttle Services. In *Advances in Intelligent Systems and Computing*, Vol. 1212 AISC. Springer, 173–185. https://doi.org/10.1007/978-3-030-50943-9_23

[39] Ralf Philipsen, Teresa Brell, and Martina Ziefle. 2018. Carriage without a Driver – User Requirements for Intelligent Autonomous Mobility Services. In *Proceedings of the International Conference on Human Factors in Transportation (AHFE '18)*, Vol. 786. Springer International Publishing, 339–350. https://doi.org/10.1007/978-3-319-93885-1

[40] SAE International. 2018. J3016-JUN2018 – Surface Vehicle Recommend Practice: Taxonomy and Definitions for Terms Related to Driving Automation Systems for On-Road Motor Vehicles. https://doi.org/10.4271/J3016_201806

[41] 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

[42] Martin Schrepp. 2018. User Experience Questionnaire Handbook. , 12 pages. https://doi.org/10.13140/rg.2.1.2815.0245

[43] 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

[44] 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

[45] Thomas Schubert, Frank Friedmann, and Holger Regenbrecht. 2001. The Experience of Presence: Factor Analytic Insights. *Presence* 10, 3 (2001), 266–281. https://doi.org/10.1162/105474601300343603

[46] Thomas Schubert, Frank Friedmann, and Holger Regenbrecht. 2016. Igroup Presence Questionnaire (IPQ). http://www.igroup.org/pq/ipq/index.php

[47] Amir Shevat. 2017. *Designing Bots*. O'Reilly Media, Inc., Sebastopol, CA. https://learning.oreilly.com/library/view/designing-bots/9781491974810/?ar

[48] 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

[49] Kyoko Sugisaki and Andreas Bleiker. 2020. Natural Language Conversation als UI - Traum oder Albtraum ? [Webinar]. https://www.germanupa.de/berufsverband-german-upa/aktuelles/webinar-natural-language-conversation-als-ui-traum-albtraum

[50] Kyoko Sugisaki and Andreas Bleiker. 2020. Usability guidelines and evaluation criteria for conversational user interfaces: A heuristic and linguistic approach. *ACM International Conference Proceeding Series* (2020), 309–319. https://doi.org/10.1145/3404983.3405505

[51] Stefano Valtolina, Barbara Rita Barricelli, and Serena Di Gaetano. 2020. Communicability of traditional interfaces VS chatbots in healthcare and smart home domains. *Behaviour and Information Technology* 39, 1 (2020), 108–132. https://doi.org/10.1080/0144929X.2019.1637025

[52] 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 arXiv:arXiv:1011.1669v3

[53] Dieter P Wallach, Lukas A Flohr, and Annika Kaltenhauser. 2020. Beyond the Buzzwords: On the Perspective of AI in UX and Vice Versa. In *Proceedings of the 1st International Conference on Artificial Intelligence in HCI, Held as Part of the 22nd HCI International Conference on Human-Computer Interaction (HCII '20)*, Helmut Degen and Lauren Reinerman-Jones (Eds.). Springer International Publishing, Cham, 146–166. https://doi.org/10.1007/978-3-030-50334-5_10

[54] Daniel Wessel, Christiane Attig, and Thomas Franke. 2019. ATI-S – An ultra-short scale for assessing affinity for technology interaction in user studies. *ACM International Conference Proceeding Series* (2019), 147–154. https://doi.org/10.1145/3340764.3340766

[55] 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

[56] Michelle X. Zhou, Gloria Mark, Jingyi Li, and Huahai Yang. 2019. Trusting virtual agents: The effect of personality. *ACM Transactions on Interactive Intelligent Systems* 9, 2-3 (2019). https://doi.org/10.1145/3232077