Don't fail me! The Level 5 Autonomous Driving Information Dilemma regarding Transparency and User Experience

Tobias Schneider
Stuttgart Media University
Stuttgart, Germany
schneidert@hdm-stuttgart.de

Joana Hois
Mercedes-Benz AG
Stuttgart, Germany
joana.hois@daimler.com

Alischa Rosenstein
Mercedes-Benz AG
Stuttgart, Germany
alischa.rosenstein@daimler.com

Sandra Metzl
Stuttgart Media University
Stuttgart, Germany
metzl@hdm-stuttgart.de

Ansgar Gerlicher
Stuttgart Media University
Stuttgart, Germany
gerlicher@hdm-stuttgart.de

Sabiha Ghellal
Stuttgart Media University
Stuttgart, Germany
ghellal@hdm-stuttgart.de

Steve Love
The Glasgow School of Art
Glasgow, Scotland
s.love@gsa.ac.uk

ABSTRACT

Autonomous vehicles can behave unexpectedly, as automated systems that rely on data-driven machine learning have shown to infer false predictions or misclassifications, e.g., due to stickers on traffic signs, and thus fail in some situations. In critical situations, system designs must guarantee safety and reliability. However, in non-critical situations, the possibility of failures resulting in unexpected behaviour should be considered, as they negatively impact the passenger's user experience and acceptance. We analyse if an interactive conversational user interface can mitigate negative experiences when interacting with imperfect artificial intelligence systems. In our quantitative interactive online survey (N=113) and comparative qualitative Wizard of Oz study (N=8), users were able to interact with an autonomous SAE level 5 driving simulation. Our findings demonstrate that increased transparency improves user experience and acceptance. Furthermore, we show that additional information in failure scenarios can lead to an information dilemma and should be implemented carefully.

CCS CONCEPTS

• **Human-centered computing** → **User studies**; *Empirical studies in visualization*; **Empirical studies in HCI**.

KEYWORDS

user experience, explainable artificial intelligence, autonomous driving

IUI '23, March 27–31, 2023, Sydney, NSW, Australia

ACM ISBN 979-8-4007-0106-1/23/03.
https://doi.org/10.1145/3581641.3584085

ACM Reference Format:
Tobias Schneider, Joana Hois, Alischa Rosenstein, Sandra Metzl, Ansgar Gerlicher, Sabiha Ghellal, and Steve Love. 2023. Don't fail me! The Level 5 Autonomous Driving Information Dilemma regarding Transparency and User Experience. In *28th International Conference on Intelligent User Interfaces (IUI '23), March 27–31, 2023, Sydney, NSW, Australia.* ACM, New York, NY, USA, 13 pages. https://doi.org/10.1145/3581641.3584085

1 INTRODUCTION

Daily interactions with systems that are based on methods of artificial intelligence (AI), particularly machine learning (ML), often lack further insights into their behaviour [Rahwan et al. 2019]. At the same time, allowing users and stakeholders to understand system behaviour better is usually recommended, if not required [High-Level Expert Group on AI 2019; P. Jonathon Phillips et al. 2020]. Systems thus have to provide some means of transparency of their behaviour. Furthermore, the system design benefits from aligning with aspects of a positive user experience (UX) to support positive user interactions. One solution to address transparency and UX in AI-based systems can be achieved by system explanations [Felzmann et al. 2020]. Designing optimal ways to provide explanations as well as their technical feasibility has thus become an active field of research [Ehsan et al. 2021; Miller 2019].

While technical improvements aim for decreasing system failures to a minimum, uncertainty in data-driven ML-based methods will always prevail. ML results can be unexpected or simply wrong. We assume that such system failures, like wrong predictions or misclassifications, cannot be eliminated entirely.

For the area of autonomous driving, studies have investigated the technical and pedestrian side of failures for external human-machine interaction (HMI) [Kuhn et al. 2020a,b; M. Faas et al. 2021], i.e., how to design the communication between vehicles and their surroundings. Yet, the communication in failure cases in autonomous driving should also address the internal HMI to avoid negative UXs for passengers. Especially for SAE level 5 [SAE On-Road Automated Vehicle Standards Committee and others 2018] autonomous driving, where vehicles do not require human control at any point of their operation and might not come with a steering wheel or pedals anymore. In general, the SAE levels define the level of driving automation ranging from level 0 with no automation but *warnings and momentary assistance* to level 5 with *fully automated driving without passenger control*. Furthermore, levels 0 to 2 are grouped as *driver support features* and require vehicle supervision at all times. The last three levels are grouped as *automated driving features*, with level 3 requiring driver takeover when needed. Levels 4 and 5 never require any takeover. However, level 4 is considered to be a high automation level since the automation is only applicable in some driving modes. In contrast, level 5 is considered full automation since the system can handle all driving modes.

Removing the possibility of control from passengers might create new challenges regarding the internal HMI communication, especially in failure cases or ambiguous situations. For example, providing explanations when the autonomous vehicle (AV) stops in front of a crosswalk due to a misclassification even though the pedestrian does not want to cross. Although we expect occurring mistakes to be non-safety-critical, they are likely to negatively impact the user's experience, increase a feeling of uncertainty and therefore diminish the acceptance of autonomous vehicles.

Furthermore, failures could even be embraced as part of the interaction design. In the context of automated driving, this has been described as *imperfection by design* [Fridman 2018], indicating that system flaws could also be seen as features to make users transparently aware of system limitations. While in some domains, when system limitations might merely be seen as an annoyance, e.g., recommender systems of online shopping sites offering advertisements of no interest, users may still continue to use those products and services. In other domains, they will shy away from using products if their reliability affects their perceived feeling of control, understanding, or even safety. Supporting these aspects by the design of AI applications is essential for user acceptance, and trust [Stanton and Jensen 2021]. However, when investigating the effects on trust, time of usage is a major influence. Recent studies have pointed out the effect of repeated interactions on trust in intelligent systems [Rossi et al. 2020; van Maris et al. 2017]. Since this paper focuses on first-time users, we focus on user acceptance and transparency in SAE level 5 autonomous driving to support a positive UX of imperfect ML systems. In particular by explicitly addressing failures with additional explanatory information.

As explanations in AI system design increase system transparency and understanding as well as user acceptance [Mcknight et al. 2011; Mohseni et al. 2018; Samek and Müller 2019], we endeavour to apply these findings particularly to the use case of failures, i.e., the inherent imperfection of an AI system in the specific context of SAE level 5 autonomous driving. Our study investigates if the expected negative UX caused by system failures can be mitigated through proactive explanations given by a conversational user interface. Moreover, users are provided with the option to receive further information through explanations upon request and the option to indicate that they would have liked to take over the driving task. It is important to mitigate the need to take over since it will not be possible in SAE level 5 autonomous driving and was not actually possible in our study design. Compared to related work discussed in the next section, we analyse how users deal with system flaws in ML systems and if their expected negative UXs can be mitigated by explanatory information. We focus on non-critical failure situations, thus fundamentally wrong driving behaviour that may even lead to road casualties, is explicitly excluded for the scope of our investigation. Our findings show that system designs face an *information dilemma*: neither a transparent system design nor an increase in transparency with additional explanatory information can improve the UX in failure situations. We discuss that a human-in-the-loop approach could thus be a solution for future designs.

2 RELATED WORK

2.1 Imperfect AI and Explainable AI

So far, some studies have addressed research questions regarding the interaction design of imperfect systems, yet clear design recommendations are still scarce. It has been shown that expectation management is crucial for designing systems that only give result predictions with a certain degree of confidence [Hase and Bansal 2020]. User acceptance increases when users are made aware of system limitations and can understand the system behaviour for them to make informed decisions [Kocielnik et al. 2019]. Explanations of system predictions positively affect user acceptance and understanding, yet explanations fall short of a positive effect in failure situations [Riveiro and Thill 2021]. Additional explanations also cause a higher awareness of users, while minimum explanations can have a negative effect [Papenmeier et al. 2019].

Several methods have been developed to make implicit AI system behaviour transparent by Explainable Artificial Intelligence (XAI) or interpretability [Adadi and Berrada 2018]. However, how such methods can immediately help to make systems transparent to end-users is yet to be determined [Samek and Müller 2019] and needs to be defined by various evaluation criteria [Mohseni et al. 2018]. For instance, in the context of human-AI interaction, further contextual aspects are relevant for a transparent interaction beyond XAI methods [Ehsan et al. 2021]. Transparency should thus be defined as a design criterion throughout the system development [Felzmann et al. 2020]. Nevertheless, the factors imposed by the human partner in the interaction context need further analysis beyond technically available explanations [Wäfler and Schmid 2020]. For instance, explanations have to be seen in the context of their complexity, as they lose their benefit when becoming incomprehensible to users [Ai et al. 2021]. For expert users and developers though, interactive visualisation have been shown to increase usability [Wang et al. 2019].

A permanent inner dialogue of a system can provide insights into system behaviour in cooperative scenarios [Pipitone and Chella 2021], yet feasibility beyond robotic (anthropomorphic) applications needs further investigation. Research has also demonstrated how users can be misled through manipulated explanations regarding the AI system's trustworthiness [Lakkaraju and Bastani 2020]. Yet, a different study, where systems that automatically translate source code into different languages, has shown how users can experience a positive co-creation situation with a system while being aware of its imperfection [Weisz et al. 2021].

For certain use cases, such as system querying for medical decisions, refinements by users during system interactions are a solution for coping with imperfect results and increasing interaction experiences and user understanding [Cai et al. 2019].

2.2 User Experience and Explainable AI in Autonomous Driving

The design of interactive systems focuses not only on providing seamless functionality but also on creating positive emotions, for example by focussing on the fulfilment of human needs, among the resulting subjective and individual experiences [Forlizzi and Battarbee 2004; Hassenzahl 2010; Hassenzahl and Tractinsky 2006; Law et al. 2009]. This UX is thereby equally important and connected to user acceptance and usability. While reliability in interactive system behaviour is essential to a good design [Nielsen 1994], ML-based systems inherently run the risk of providing wrong predictions, violating best practices of interactive system design. In this regard, supporting user acceptance, usability and thus a positive UX is essential to automated systems.

This also applies in the context of autonomous driving, which is based on automated systems. Studies show that transparent communication is essential for trust and user acceptance in AVs [Abraham et al. 2016; Ha et al. 2020; Iclodean et al. 2020; Koo et al. 2015]. Different uni- and multimodal feedback modalities for making vehicles and AVs more transparent, hence designing for XAI communication, have been researched, using auditory, vibrotactile, visual, textual, light, augmented reality and on-device (smartphone) cues. Such design methods comprise general driver warning (see for example [Ho et al. 2005; Politis et al. 2013, 2015b, 2014]), takeover requests (see for example [Borojeni et al. 2016; Geitner et al. 2019; Gold et al. 2013; Huang et al. 2019; Melcher et al. 2015; Petermeijer et al. 2017; Politis et al. 2015a, 2017; Salminen et al. 2019; Telpaz et al. 2017; Walch et al. 2015; Zeeb et al. 2015]), or uncertainty communication (see for example [Beller et al. 2013; Faltaous et al. 2018; Kunze et al. 2018; Noah et al. 2017; Seppelt and Lee 2007]). In particular, it has also been shown that there is a demand by users to receive additional information in case of unexpected behaviour of a vehicle [Wiegand et al. 2020].

XAI pursues the goal of the clarification and enhancement of the AI system interaction benefitting the end users, in our case the passenger. Therefore, it is crucial to take into account that passengers require or prefer various explanations based on the situation or the desired result [Miller 2019], which is consistent with the subjective character of UX. However, while many studies looked at different uni- and multimodal feedback modalities in autonomous driving, few have focussed on the combination of UX and XAI. Large et al. [2019] were able to show that transparent communication can increase the UX. We were able to confirm this effect in another study [Schneider et al. 2021a], as were Detjen et al. [2021] and Dandekar et al. [2022]. Furthermore, we were able to show that additional explanatory information in SAE level 5 autonomous driving not only prevents negative experiences during or after the ride but can increase the perceived feeling of safety and control even though users cannot interfere with the driving task [Schneider et al. 2021b,c]. We demonstrated that the positive effect of additional information on UX was fully mitigated by the loss of subjective feeling of control, which describes the users' impression of having sufficient control over the system and the situation as a whole [Schneider et al. 2021b]. Moreover, users also show individual preferences regarding the amount of information received in situations and their desire to take over the driving task [Park et al. 2020].

While explanations for transparent system design in autonomous driving have been applied at various system levels and different user groups in mind [Omeiza et al. 2021], we particularly investigate the handling of (non-critical) system failures with explanations in autonomous driving with different feedback modalities in the following.

3 EXPERIMENTAL SETUP

3.1 Prototype

In the context of SAE level 5 autonomous driving, where no human attention or interaction is required, we developed an interactive conversational user interface using the Telegram messenger to provide passengers with live feedback about driving situations and the AV's perception of those on a central display. In a prior study, we showed that text is a pragmatic way to communicate information in autonomous driving [Schneider et al. 2021a]. To avoid information overflow and allow for approximating the optimal level of detail, the system displays short text messages about the AV's perception of driving situations to explain its reactions. For example, the AV brakes in front of a crosswalk and displays the message *"I spotted a person on the side of the road"*. It furthermore offers passengers the possibility to receive more detailed information on demand. This includes a longer textual description of the situation, visualisations and object highlighting of what the vehicle recognised in a given driving scene (see Figure 1).

Our prototype presents users a 2.5km long route (see Figure 2) with six different driving situations (see Table 1). The first three are regular situations where nothing unusual happens. The last three are failure situations where the AV did not interpret a situation correctly and therefore performed a wrong action. These are examples of potential difficulties for scene understanding and misclassifications of the situation or the intent of other road users by an autonomous driving system [Janai et al. 2020]. The car was driving with a 50km/h limit for the first three situations and a 30km/h limit for the last three situations, according to the speed limit given by the public road signs for the route.

3.2 Experiment Design

We conducted a mixed-method within- and between-subjects study as an online- and on-site experiment. Due to the health and safety legislation during the pandemic, the number of on-site participants was limited. The primary purpose of the online study was to collect quantitative data in the form of questionnaires. However, participants were also given a text field for comments to leave qualitative feedback. The main purpose of the on-site experiment was to collect qualitative data in the form of think-aloud statements [Lewis 1982] and participant observation. A summary of the experimental design can be seen in Figure 4.

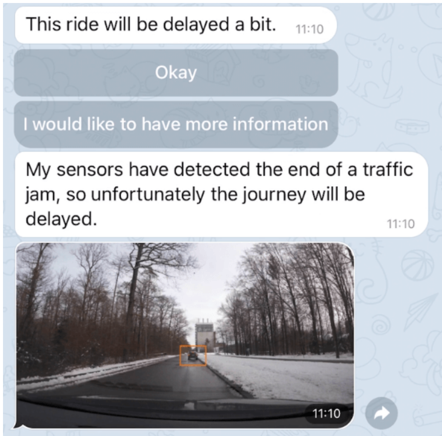


Figure 1: Screenshot of the conversational user interface and the traffic jam fail situation. The user selected the second option, and therefore a more descriptive text and image were displayed.

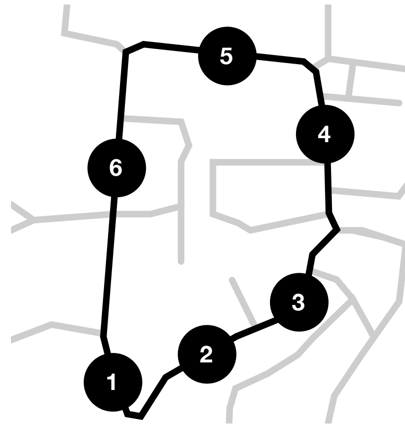


Figure 2: Illustration of the simulated autonomous drive route with the six different situations.

Table 1: At six locations of the route during the simulated autonomous driving setup, the conversational system shows information about the current driving situation and provides additional information to passengers upon request, i.e., passengers were able to select the option "I would like to have more information".

No.	Situation	Displayed Description	Explanation upon Request
(1)	The AV correctly detects a crosswalk where no pedestrian is crossing.	"I have detected a pedestrian crossing."	An image of the crosswalk with a highlight is shown.
(2)	The AV correctly detects a green traffic light.	"I have detected a traffic light."	An image of the traffic light with a highlight is shown.
(3)	The AV correctly detects a green traffic light.	"I have detected a traffic light."	An image of the traffic light with a highlight is shown.
(4)	The AV misinterprets a person standing next to a crosswalk as someone planning to cross the road. The person has no intent to do so, yet, the AV does not continue the drive for 10 seconds.	"I detected a person at the roadside."	An image of the crosswalk with the pedestrian highlighted is shown. "I have detected a pedestrian at the side of the road who might plan to walk onto the road. I'll wait until the pedestrian has crossed the road."
(5)	The AV misinterprets multiple parked cars as the end of a traffic jam and comes to a stop behind them. It continues the drive after 12 seconds and passes the parked cars.	"This ride will be delayed a bit."	An image of the crosswalk with the pedestrian highlighted is shown. "My sensors have detected the end of a traffic jam, so unfortunately the journey will be delayed."
(6)	The AV misinterprets an advertisement poster showing an advert campaign with a stop sign as a regular stop sign and comes to a halt. It continues the drive after the complete stop.	"I have spotted a stop sign."	An image of the highlighted advertisement poster is shown.

3.2.1 Online Experiment. For the online experiment, we developed an interactive online questionnaire in which the different driving situations and text responses were shown as a video (see Figure 3a and Appendix A). Participants were presented with a pre-recorded autonomous drive, which took 4:50 minutes. After every situation, they were asked if they hypothetically would have preferred to take over the driving task. However, they had no option to actually take over control, as we focused on SAE level 5. Furthermore, the participants of the online experiment were divided into two groups (control and experimental). In contrast to the control group, the participants of the experimental group were asked after every situation if they wanted to request more information. If so, they were directed to an additional page of the online-questionnaire where textual information and an image of the situation were displayed. We decided on this approach to be able to evaluate the wish for additional information and the effect of the provided information on the UX in contrast to having no option for additional information. Since the participants were able to request additional information through the user interface and indicate their preference to take over the driving task, behavioural

(a) Online Visualisation

(b) On-site Setup

Figure 3: (a) Screenshot of the online-experiment. (b) Photos of the onsite-experiment. A curtain separated driver and passenger. A tablet with the conversational user interface was mounted in front of the passenger.

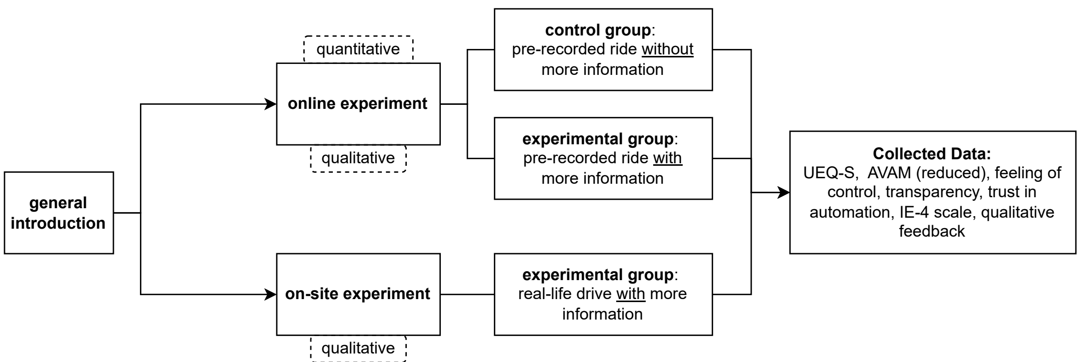


Figure 4: Study setup.

data could be collected in addition to the attitudinal data from the questionnaires.

3.2.2 On-Site Experiment. For the on-site experiment, we chose to investigate further the experimental condition in which participants were able to request feedback from the system. By doing so, deeper insights explaining the collected attitudinal and behavioural data from the quantitative study could be derived. We created an experiment adapted from a Wizard of Oz study [Kelley 1984] with an electric vehicle. The route and situations were the same as in the online experiment. We installed a black curtain to separate the passenger from the driver to emulate an SAE level 5 autonomous drive. Participants were generally introduced to the experiment and informed that a driver is present. However, they could not see or interact with the driver due to the curtain. We installed a tablet in front of the passenger where the conversational user interface was displayed and could be interacted with (see Figure 3b). Unlike in the online experiment, all participants could request more information. Furthermore, participants were asked to verbalise their thoughts during the experiment and to state if, in any situation, they would prefer to take over the driving task. These participants also filled out the same questionnaires as the online participants.

3.2.3 Questionnaires. After experiencing the simulated autonomous ride (on-site or online), the participants filled out multiple questionnaires:

1) The User Experience Questionnaire - Short (UEQ-S)[Schrepp et al. 2017] to measure their user experience during the ride: *How would you rate the ride, especially in terms of vehicle behaviour and communication?*
2) A reduced version of the Autonomous Vehicle Acceptance Model Questionnaire (AVAM) [Hewitt et al. 2019], focussing on the variables *attitude towards (using) technology*, *anxiety*, *behavioural intention (to use the vehicle)* and *perceived safety*.
3) A self-defined 3-item, 7-point Likert scale regarding the feeling of control, see Table 2.
4) A self-defined 2-item, 7-point Likert scale regarding the transparency of the system, see Table 2.
5) A reduced version of the questionnaire by Körber[Körber 2018] that measures trust in automation, focussing on the variables *understanding/predictability* and *reliability/competence*.
6) The IE-4 scale[Kovaleva 2012] as a control variable to measure the internal and external locus of control.

3.3 Participant Groups

3.3.1 Online Experiment. Overall, 113 participants (41 female) were individually taking part in the online experiment (70 students, 37 employees, 6 other). 44 of them were in the control group that was not provided with the option to request more information. Their average age was 27.7 (SD=10.13). The remaining 69 participants were in the experimental group that was provided with the option

Table 2: Questionnaire items for the perceived feeling of control and transparency. The AVAM questionnaire wording was used to formulate the items for control, i.e., participants received comparable formulations when giving feedback on all items.

Perceived Feeling of Control
I would have adequate control over the information given by the vehicle
I would have adequate control to get the information that I need about the ride
I have the feeling of being in control during the ride
Transparency
The system is transparent
My interaction with the vehicle would be clear and understandable (AVAM item 5)

to request more information. Their average age was 27.3 (SD=9.60). We decided on a larger number of participants for the experimental condition since participants were given an additional option to interact with the system, i.e. requesting more information. By doing so we ensured a sufficient number of participants for all possible combinations of user choices, i.e. requesting more information and requesting to take over the driving task hypothetically.

The majority of participants primarily travel by car with 55.8%, 24.8% travel by public transport, 11.5% by foot, 6.2% by bicycle and 1.7% use other main ways of transportation. Based on their prior experiences with autonomous systems, participants are distributed in: 36.3% had no prior experiences, 39.8% had experiences with driving assistance such as cruise control, 16.8% had experiences with semi-automatic driving systems such as lane assists, 5.3% had experiences with highly automated driving systems such as highway and takeover driving assistants, and 1.8% had experiences with fully AVs in research contexts.

3.3.2 On-Site Experiment. For the on-site experiment, 8 participants (1 female) participated individually (6 students, 1 employee, 1 other). Their average age was 24.5 (SD=1.60). With 4 of them mainly travelling by public transport, 2 by foot, 1 by car and 1 by bicycle. 3 of them had no prior experience with autonomous systems, 2 had experience with driving assistance, 2 with semi-automatic driving systems and 1 with highly automated driving systems.

3.4 Hypotheses

The assumed hypotheses for the experimental setup are:

- H_{1a} *Hypothesis$_{1a}$*: A system failure increases the users' need to take over the driving task.
- H_{1b} *Hypothesis$_{1b}$*: Additional information upon request (higher level of transparency) reduce the users' need to take over the driving task when a system failure occurs.
- H_{2a} *Hypothesis$_{2a}$*: Additional information upon request (higher level of transparency) increase the user experience.
- H_{2b} *Hypothesis$_{2b}$*: Additional information upon request (higher level of transparency) increase the users' system acceptance.
- H_{2c} *Hypothesis$_{2c}$*: Additional information upon request (higher level of transparency) increase the users' subjective feeling of control.
- H_{3a} *Hypothesis$_{3a}$*: Users demand additional information when experiencing system failure.
- H_{3b} *Hypothesis$_{3b}$*: Additional information upon request (higher level of transparency) can mitigate the negative effects of system failures on user experience.

4 RESULTS

4.1 User Interactions During the Experiment

Figure 5 illustrates which decisions and feedback users decided to select based on the experimental situations shown in Figure 4. On average, more additional information is requested by users when they encounter a situation that was misinterpreted by the system in situations 4-6. Also, their desire to take over in these situations is significantly higher (even though not possible due to SAE level 5). As for the control group, which did not have the option to request additional information, their desire to take over is twice as high.

4.2 Quantitative Results

To verify our hypotheses, a number of statistical analyses were performed:

(1) exploratory, descriptive analyses were performed on the questionnaire results
(2) Cronbach's Alpha was calculated for all variables to ensure an adequate internal consistency; Table 3 reports means, standard deviations, and internal consistency
(3) intercorrelations of the studied variables, i.e., transparency, dimensions of the AVAM and the UEQ-S were calculated (see Table 4)
(4) in-between and between-subject comparisons using the t-test and the ANOVA were performed. Since the direction of effects was predicted, all results report a one-tailed significance level of $p<.05$

Table 3: Descriptive measures of used variables.

Variables	Min-Max	Mean (SD)	a
(1) Transparency	1-7	3.16 (1.33)	.71
(2) PQ of the UX	1-7	3.33 (1.42)	.81
(3) HQ of the UX	1-7	3.98 (1.49)	.87
(4) Perc. feeling of control	1-7	3.63 (1.33)	.80
(5) AVAM - Effort	1-7	2.50 (1.40)	.68
(6) AVAM - Attitude	1-7	3.65 (1.53)	.79
(7) AVAM - Anxiety	1-7	4.71 (1.44)	.71
(8) AVAM - Intention to use	1-7	3.63 (1.60)	.79
(9) AVAM - Safety	1-7	3.33 (1.42)	.75

Due to a lack of comparability, the data from the on-site experiment was not included in the statistical analysis. However, it supports the findings from the online study, so the results seem to be transferable to a real-life driving situation. Regarding hypotheses H_{1a}-H_{1b}, the results show a significant difference in take over requests between failure and non-failure situations

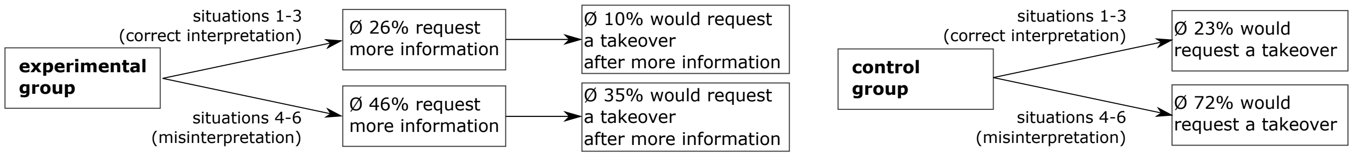


Figure 5: Information requests and takeover decisions by situation and participant group.

Table 4: Intercorrelations between transparency and the dependent variables.

Variable	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
(1) Transparency								
(2) Pragmatic Quality of the UX	.39**							
(3) Hedonic Quality of the UX	.30**	.46**						
(4) Perc. feeling of control	.58**	.50**	.33**					
(5) AVAM - Effort	.49**	.38**	.22**	.40**				
(6) AVAM - Attitude	.34**	.65**	.48**	.57**	.33**			
(7) AVAM - Anxiety	-.24**	-.25**	-.04	-.35**	-.17*	-.26**		
(8) AVAM - Intention to use	.35**	.54**	.26**	.53**	.17*	.65**	-.48**	
(9) AVAM - Safety	.29**	.36**	.10	.47**	.18*	.49**	-.59**	.67**

(p<.01, t= -11.24). Hypothesis H_{1a} could therefore be confirmed. However, providing additional information upon request did not influence the preference to take over the driving task in a failure situation (p>.05, t= -.15). Hypothesis H_{1b} was therefore rejected.

In order to verify the hypotheses H_{2a}-H_{2c}, the intercorrelations between transparency and the dependent variables pragmatic quality (PQ) of the UX (r=.39, p<.01), hedonic quality (HQ) of the UX (r=.30, p<.01), all facets of the system acceptance (effort: r=.49, p<.01; attitude: r=.34, p<.01; anxiety: r=-.24, p<.01; intention to use: r=.35, p<.01; safety: r=.29, p<.01) and the perceived feeling of control (r=.58, p<.01) were considered. As seen in Table 4, significant correlations between the variables in the expected direction were found in all cases.

We tested for direct and indirect effects using the Sobel test (mediation analysis) to examine further the relationship between transparency, the perceived feeling of control, and the UX. We aimed to develop a deeper understanding of the mechanisms of action of transparency in the context of UX.

Regarding the PQ of the UX, a significant indirect effect of transparency mediated by the subjective feeling of control was found (p<.01), with the standardised indirect effect being .39. However, the direct effect of transparency on PQ was no longer significant when taking the direct effect of subjective feeling of control into account (p>.05). Therefore, the relationship between transparency and the PQ of the UX is fully mediated by the subjective feeling of control.

Accordingly, the indirect effect of transparency on the HQ of the UX was examined. Again, a significant indirect effect of transparency on HQ, mediated by the subjective feeling of control, was found (p<.01), with the standardised indirect effect being .58. When controlling for the direct effect of subjective feeling of control, the direct effect of transparency on HQ was no longer significant (p>.05). Therefore, the relationship between transparency and HQ is fully mediated by the subjective feeling of control.

However, there was no significant difference in the participants' transparency level between the experimental groups. Additional information upon request about the system failure did not increase the subjective transparency level (p>.05, t= 1.63). Accordingly, there was no significant difference between the participants who received additional information and those who did not, regarding the dependent variables PQ of the UX (p>.05, F=.64), HQ of the UX (p>.05, F= 1.43) and all facets of the system acceptance (effort: p>.05, F= .01; attitude: p>.05, F=.10; anxiety: p>.05, F= .05; intention to use: p>.05, F= .27; safety: p>.05, F= .01). Only in the case of the subjective feeling of control a significant increase for the participants who received additional information compared to those who did not could be shown (p<.05, F= 3.51). The hypotheses H_{2a}-H_{2c} can therefore be only partially confirmed regarding the effect of the subjective level of transparency.

Moreover, the need for additional information when confronted with a system failure was examined. It could be shown that participants do request additional information more often when faced with a system failure compared to normal system behaviour (p<.05, t= -1.87). Hypothesis H_{3a} could therefore be confirmed. Furthermore, a mitigating effect of additional information upon request on the negative effects of system failures on UX and acceptance was expected. The results of the ANOVA, however, did not confirm an interaction effect of additional information on PQ of the UX (p>.05, F= .90), HQ of the UX (p>.05, F= .43). Thus, hypothesis H_{3b} was rejected.

4.3 Qualitative Results

The online study contained an optional field for concluding feedback comments from participants. 34 participants provided additional comments. Together with the 8 on-site participants, we conducted a two-annotator-based categorisation in the context of our hypotheses. Table 5 shows an overview of their collectively expressed comments. Several comments addressed transparency of the system, e.g., *"I regarded transparency and communication of the*

Table 5: Numbers of on-site and online participants expressing their opinions on the summarised category statements.

Statement category	# On-site	# Online
Transparency of the system was recognised by participants.	6	3
Transparency of the system was unnecessary.	-	2
Transparency of the system was not enough.	-	3
Participants would have preferred to take over the driving task.	3	3
Participants doubt that the system would be accepted by customers.	1	6
Participants are satisfied with the driving style.	4	-

system as being convincing overall." (online participant), *"I miss an exact visualisation of the vehicle's internal representation of what it recognises."* (online participant), *"Alright, the pedestrian was recognised but does not cross the street. [The vehicle] can continue driving."* (on-site participant). Some comments also addressed a preference to take over the driving task, e.g., *"I would take over in this situation, as the vehicle breaks unnecessarily."* (on-site participant), *"I do not think I would be interested in asking for more information. If I could not understand the vehicle manoeuvring, I would take over."* (online participant).

As on-site participants were asked to express their opinion verbally, more feedback was collected from those participants. Specific comments about transparency and user acceptance were less directly expressed. Overall, a slightly positive trend towards the recognised transparency of the system can be seen in both experiments, more often, however, for the on-site experiment. And clearly, only on-site participants expressed their opinion on the actual driving style (e.g., they were satisfied that the turn light was used). This qualitative feedback by participants is far from an exhaustive interpretation and analysis, yet the overview indicates that users tend to experience the system as being transparent.

5 DISCUSSION

Our experiment shows that passengers of an SAE level 5 AV would prefer to take over the driving task in some situations, particularly in failure situations, and their need to take over control is expressed. To foster autonomous driving, this need for control can be mitigated by a design with additional explanatory information to increase transparency. We demonstrated this in a prior study using live as well as additional retrospective information about the AV's actions, which significantly increased the perceived feeling of control for passengers even though they were not able to take over the driving task [Schneider et al. 2021b]. This current study further supports this effect, showing that transparency increases the perceived feeling of control, which increases both facets (pragmatic and hedonic) of the UX.

The option to request even more system information in failure situations, however, did neither mitigate the need to take over control nor increase the transparency or UX. This outcome could be caused by the requested information only pointing out a failure situation in even more detail.

Regarding the feeling of control, 6 testers (4 on-site) expressed that they would have liked to take over control in one or more of the failure situations ("*I would definitely intervene here [traffic jam fail] now. Or at least say that it [the AV] can move on.*", "*Now, I would intervene and continue driving. Because the person does not walk onto the crosswalk.*").

Statements of the participants further confirmed the usefulness of transparency. 9 participants (7 on-site) positively commented on the transparency of the system ("*It was good that it [the conversational user interface] informed me in this situation.*"). 3 online participants wished for more transparency ("*The vehicle would provide more transparency if the navigation, i.e., map and route, were displayed while driving.*") and 2 said that the transparency was too much and/or unnecessary ("*I miss the option: Interaction is unnecessary. [...] I would be grateful if the vehicle did not interact with me [...]*").

The lack of positive effects from additional explanatory information by the system in failure cases was also shown in a study with factual and counterfactual explanations [Riveiro and Thill 2021], where no significant increase in user acceptance, trust, or understanding was found for counterfactual explanations of classification failures by a text classification application. We interpret those results in a similar way to our findings that failure situations fall short on positive UX because the explanations actually point out incorrect system behaviour. This was also confirmed by 6 participants (3 on-site) who criticised the three failure situations of the AV ("*That was a somehow bad estimation, as the car is parked.*", "*It would definitely upset me over time if something like this happens more often.*").

Although participants experienced failure situations, no significant difference between their attitudes towards autonomous driving was recognisable before and after the experiment – neither for the group with nor for the group without the option to request more information. There might be several reasons for this: Firstly, our study was predominately conducted online with videos and no real-world experience. Secondly, it was a one-time experience, and a long-term exposure might change their opinion. Thirdly, it might be caused by our study population being generally more positive towards autonomous driving. 4 participants (3 on-site) showed tolerance towards failures ("*That's not a stop sign (laughs).*", "*That [traffic jam fail] is, of course, impractical if something like that is not recognised. But it was solved relatively quickly.*").

Furthermore, participants increasingly requested more information in failure situations. This implies that they did not understand the AV's behaviour in these situations or that they recognised the failure. In turn, a user request for more information may indicate that the AV is acting incorrectly. 3 testers (1 on-site) wished for a way to interact with the AV or to give feedback ("*I would like to have the opportunity to rate the autonomous driving system.*", "*In the situation with the pedestrian near the crosswalk, I* *would have preferred to have a button to continue – not interfere directly with the control, but to give my recommendation.*"). From a general design point of view, if an AV did provide the option for additional information upon request, it would be likely that an increase in requests in certain situations point out failure situations.

6 LIMITATIONS

One limitation of our study is that we performed most of the experiments online due to the pandemic situation. Our interactive driving videos thus lack the physical feeling of an actual driving situation. However, correlations of participant behaviour in driving simulator studies have been shown to be applicable to real-world behaviour [Mullen et al. 2011], in particular for non-critical situations.

The on-site study, with its focus on qualitative feedback, was carried out as an adapted Wizard of Oz study design. Participants were informed that they are driven by a human driver, even if they cannot see them. Their verbal expressions, however, do not indicate that this affected their experience of an AV simulation and the conversational user interface. Users expressed their impression of the system and the interface, not the real driver.

Regarding the participant population, demographic limitations are evident. More than half of the participants were university students. The average age of participants represents a younger population. A broader range of participants would have been preferred, especially elderly passengers. This would have allowed us to generalise our research findings more broadly.

7 CONTRIBUTION & CONCLUSION

The design of transparent communication in autonomous driving faces a challenge in case of AV failure situations. While the design of semi-automated driving, as well as the external communication with pedestrians, has to address the challenges of over-trust [Fridman 2018; M. Faas et al. 2021], the design of internal communication in SAE level 5 autonomous driving has to address the perceived feeling of control and safety, understanding and overall acceptance of the passengers. And while we have formerly shown that in regular situations, internal transparent communication increases the factors mentioned above, as well as the UX [Schneider et al. 2021b], different factors arise with failure situations. Communicating system limitations might not be a solution if passengers do not have the possibility to take over control in SAE level 5 autonomous driving scenarios. Since the system is not aware that it is making a mistake, providing more *incorrect* information does not support the passengers. In contrast, it has a negative effect on the UX. It is a fine line between gaining and losing the passengers' acceptance.

Thus, regarding the system design of an AV, this creates an information dilemma. Transparent communication contributes to positive UX and user acceptance but is not helpful if it reveals the limitations of the system by explaining behaviour that is unintended or unexpected from the passenger's point of view. As the system does not have any information about its failure, a feedback loop between the human and the AV could be used to mitigate the possible negative effects of said limitations. This might be especially true when one considers the way transparency works in the context of UX. Since it could be shown that the effect of transparency on UX is fully mediated by the subjective feeling of control, enabling a feedback loop to actively involve the human seems beneficial. For instance, an increase in requests for more information can be seen as an indication of a failure situation or at least a situation in which the system behaviour is seen as unintended or not understood by the passenger. Therefore, an increase in requests can be used as an indicator for designing an interface that preventatively provides additional information to the passengers.

Another idea could be to allow the passenger to give feedback to the system that an explanation was not helpful or that its behaviour was not understood, not correct or not as expected. Such a human-in-the-loop design [5000.59 1998] could help mitigate failure situations as well as address system design limitations. Alternatively, users could be offered the possibility to contact a human operator for teleoperation [Kettwich et al. 2021], specifically regarding failure situations.

To conclude, we contribute the following insights and design suggestions:

1) In regular driving situations, internal transparent communication contributes to positive UX and user acceptance. It has its limits in failure situations.
2) The effect of transparency on the UX is fully mediated by the subjective feeling of control.
3) SAE level 5 autonomous driving could use a human-in-the-loop concept for passengers to mitigate potential UX, control and safety issues.
4) The request for more information is an indication that something feels wrong or unexpected to the passenger.
5) The human-in-the-loop design can thus be designed proactively based on the number of user requests.
6) Based on the number of user requests, offering to contact a human operator for teleoperation could be considered for failure situations.

We support further research regarding the area of failure communication in SAE level 5 autonomous driving and its implication on the passengers' perceived feeling of control and safety towards the system and their UX.

ACKNOWLEDGMENTS

This work was carried out as part of the FlexCAR project and partially funded by the German Federal Ministry of Education and Research (funding number: 02P18Q647). It was executed in cooperation with the research campus Arena2036, the Stuttgart Media University and the Mercedes-Benz AG. The conversational user interface programming and the recording of the driving scenes were carried out by students of the Stuttgart Media University based on our specifications (in alphabetical order): T. Altrichter, M. Audring, S. Haas, J. Heinzelmann, O. Maicher, E. March, C. Walter.

REFERENCES

DoD Directive 5000.59. 1998. DoD Modeling and Simulation (M&S) Glossary. (1998).
Hillary Abraham, Chaiwoo Lee, Samantha Brady, Craig Fitzgerald, Bruce Mehler, Bryan Reimer, and Joseph F Coughlin. 2016. Autonomous vehicles, trust, and driving

alternatives: A survey of consumer preferences. *Massachusetts Inst. Technol, AgeLab, Cambridge* 1, 16 (2016), 2018–12.

Amina Adadi and Mohammed Berrada. 2018. Peeking Inside the Black-Box: A Survey on Explainable Artificial Intelligence (XAI). *IEEE Access* 6 (2018), 52138–52160. https://doi.org/10.1109/ACCESS.2018.2870052

Lun Ai, Stephen H. Muggleton, Céline Hocquette, Mark Gromowski, and Ute Schmid. 2021. Beneficial and harmful explanatory machine learning. *Machine Learning* 110 (2021), 695–721. Issue 4. https://doi.org/10.1007/s10994-020-05941-0

Johannes Beller, Matthias Heesen, and Mark Vollrath. 2013. Improving the driver–automation interaction: An approach using automation uncertainty. *Human factors* 55, 6 (2013), 1130–1141.

Shadan Sadeghian Borojeni, Lewis Chuang, Wilko Heuten, and Susanne Boll. 2016. Assisting drivers with ambient take-over requests in highly automated driving. In *Proceedings of the 8th international conference on automotive user interfaces and interactive vehicular applications*. 237–244.

Carrie J. Cai, Emily Reif, Narayan Hegde, Jason Hipp, Been Kim, Daniel Smilkov, Martin Wattenberg, Fernanda Viegas, Greg S. Corrado, Martin C. Stumpe, and Michael Terry. 2019. Human-Centered Tools for Coping with Imperfect Algorithms During Medical Decision-Making. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (Glasgow, Scotland Uk) *(CHI '19)*. Association for Computing Machinery, New York, NY, USA, 1–14. https://doi.org/10.1145/3290605.3300234

Aditya Dandekar, Lesley-Ann Mathis, Melanie Berger, and Bastian Pfleging. 2022. How to Display Vehicle Information to Users of Automated Vehicles When Conducting Non-Driving-Related Activities. *Proceedings of the ACM on Human-Computer Interaction* 6, MHCI (2022), 1–22.

Henrik Detjen, Maurizio Salini, Jan Kronenberger, Stefan Geisler, and Stefan Schneegass. 2021. Towards Transparent Behavior of Automated Vehicles: Design and Evaluation of HUD Concepts to Support System Predictability Through Motion Intent Communication. In *Proceedings of the 23rd International Conference on Mobile Human-Computer Interaction*. 1–12.

Upol Ehsan, Q. Vera Liao, Michael Muller, Mark O. Riedl, and Justin D. Weisz. 2021. Expanding Explainability: Towards Social Transparency in AI Systems. 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 82, 19 pages. https://doi.org/10.1145/3411764.3445188

Sarah Faltaous, Martin Baumann, Stefan Schneegass, and Lewis L Chuang. 2018. Design guidelines for reliability communication in autonomous vehicles. In *Proceedings of the 10th international conference on automotive user interfaces and interactive vehicular applications*. 258–267.

Heike Felzmann, Eduard Fosch-Villaronga, Christoph Lutz, and Aurelia Tamò-Larrieux. 2020. Towards Transparency by Design for Artificial Intelligence. *Science and Engineering Ethics* 26 (2020), 3333–3361. Issue 6. https://doi.org/10.1007/s11948-020-00276-4

Jodi Forlizzi and Katja Battarbee. 2004. Understanding Experience in Interactive Systems. In *Proceedings of the 5th Conference on Designing Interactive Systems: Processes, Practices, Methods, and Techniques* (Cambridge, MA, USA) *(DIS '04)*. Association for Computing Machinery, New York, NY, USA, 261–268. https://doi.org/10.1145/1013115.1013152

Lex Fridman. 2018. Human-Centered Autonomous Vehicle Systems: Principles of Effective Shared Autonomy. *ArXiv* abs/1810.01835 (2018).

Claudia Geitner, Francesco Biondi, Lee Skrypchuk, Paul Jennings, and Stewart Birrell. 2019. The comparison of auditory, tactile, and multimodal warnings for the effective communication of unexpected events during an automated driving scenario. *Transportation research part F: traffic psychology and behaviour* 65 (2019), 23–33.

Christian Gold, Daniel Damböck, Lutz Lorenz, and Klaus Bengler. 2013. "Take over!" How long does it take to get the driver back into the loop?. In *Proceedings of the human factors and ergonomics society annual meeting*, Vol. 57. Sage Publications Sage CA: Los Angeles, CA, 1938–1942.

Taehyun Ha, Sangyeon Kim, Donghak Seo, and Sangwon Lee. 2020. Effects of explanation types and perceived risk on trust in autonomous vehicles. *Transportation research part F: traffic psychology and behaviour* 73 (2020), 271–280.

Peter Hase and Mohit Bansal. 2020. Evaluating Explainable AI: Which Algorithmic Explanations Help Users Predict Model Behavior?. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*. Association for Computational Linguistics, Online, 5540–5552. https://doi.org/10.18653/v1/2020.acl-main.491

Marc Hassenzahl. 2010. Experience design: Technology for all the right reasons. *Synthesis lectures on human-centered informatics* 3, 1 (2010), 1–95.

Marc Hassenzahl and Noam Tractinsky. 2006. User experience-a research agenda. *Behaviour & information technology* 25, 2 (2006), 91–97.

Charlie Hewitt, Ioannis Politis, Theocharis Amanatidis, and Advait Sarkar. 2019. Assessing public perception of self-driving cars: The autonomous vehicle acceptance model. In *Proceedings of the 24th international conference on intelligent user interfaces*. 518–527.

High-Level Expert Group on AI. 2019. *Ethics guidelines for trustworthy AI*. Technical Report. European Commission. Retrieved May, 2021 from https://ec.europa.eu/digital-single-market/en/news/ethics-guidelines-trustworthy-ai

Cristy Ho, Hong Z Tan, and Charles Spence. 2005. Using spatial vibrotactile cues to direct visual attention in driving scenes. *Transportation Research Part F: Traffic Psychology and Behaviour* 8, 6 (2005), 397–412.

Gaojian Huang, Clayton Steele, Xinrui Zhang, and Brandon J Pitts. 2019. Multimodal cue combinations: a possible approach to designing in-vehicle takeover requests for semi-autonomous driving. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, Vol. 63. SAGE Publications Sage CA: Los Angeles, CA, 1739–1743.

Calin Iclodean, Nicolae Cordos, and Bogdan Ovidiu Varga. 2020. Autonomous shuttle bus for public transportation: A review. *Energies* 13, 11 (2020), 2917.

Joel Janai, Fatma Guney, Aseem Behl, and Andreas Geiger. 2020. *Computer Vision for Autonomous Vehicles: Problems, Datasets and State-of-the-Art*. Now Publishers Inc.

John F Kelley. 1984. An iterative design methodology for user-friendly natural language office information applications. *ACM Transactions on Information Systems (TOIS)* 2, 1 (1984), 26–41.

Carmen Kettwich, Andreas Schrank, and Michael Oehl. 2021. Teleoperation of Highly Automated Vehicles in Public Transport: User-Centered Design of a Human-Machine Interface for Remote-Operation and Its Expert Usability Evaluation. *Multimodal Technologies and Interaction* 5, 5 (2021), 26.

Rafal Kocielnik, Saleema Amershi, and Paul Bennett. 2019. Will You Accept an Imperfect AI? Exploring Designs for Adjusting End-user Expectations of AI Systems. In *CHI 2019*. ACM. https://www.microsoft.com/en-us/research/publication/will-you-accept-an-imperfect-ai-exploring-designs-for-adjusting-end-user-expectations-of-ai-systems/

Jeamin Koo, Jungsuk Kwac, Wendy Ju, Martin Steinert, Larry Leifer, and Clifford Nass. 2015. Why did my car just do that? Explaining semi-autonomous driving actions to improve driver understanding, trust, and performance. *International Journal on Interactive Design and Manufacturing (IJIDeM)* 9, 4 (2015), 269–275.

Moritz Körber. 2018. Theoretical considerations and development of a questionnaire to measure trust in automation. In *Congress of the International Ergonomics Association*. Springer, 13–30.

Anastassiya Kovaleva. 2012. *The IE-4: Construction and validation of a short scale for the assessment of locus of control*. Vol. 9. DEU.

Christopher B Kuhn, Markus Hofbauer, Goran Petrovic, and Eckehard Steinbach. 2020a. Introspective black box failure prediction for autonomous driving. In *2020 IEEE Intelligent Vehicles Symposium (IV)*. IEEE, 1907–1913.

Christopher B Kuhn, Markus Hofbauer, Goran Petrovic, and Eckehard Steinbach. 2020b. Introspective Failure Prediction for Autonomous Driving Using Late Fusion of State and Camera Information. *IEEE Transactions on Intelligent Transportation Systems* (2020).

Alexander Kunze, Stephen J Summerskill, Russell Marshall, and Ashleigh J Filtness. 2018. Augmented reality displays for communicating uncertainty information in automated driving. In *Proceedings of the 10th international conference on automotive user interfaces and interactive vehicular applications*. 164–175.

Himabindu Lakkaraju and Osbert Bastani. 2020. "How Do I Fool You?": Manipulating User Trust via Misleading Black Box Explanations. In *Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society* (New York, NY, USA) *(AIES '20)*. Association for Computing Machinery, New York, NY, USA, 79–85. https://doi.org/10.1145/3375627.3375833

David R Large, Kyle Harrington, Gary Burnett, Jacob Luton, Peter Thomas, and Pete Bennett. 2019. To please in a pod: employing an anthropomorphic agent-interlocutor to enhance trust and user experience in an autonomous, self-driving vehicle. In *Proceedings of the 11th international conference on automotive user interfaces and interactive vehicular applications*. 49–59.

Effie Lai-Chong Law, Virpi Roto, Marc Hassenzahl, Arnold POS Vermeeren, and Joke Kort. 2009. Understanding, scoping and defining user experience: a survey approach. In *Proceedings of the SIGCHI conference on human factors in computing systems*. 719–728.

Clayton Lewis. 1982. *Using the" thinking-aloud" method in cognitive interface design*. IBM TJ Watson Research Center Yorktown Heights.

Stefanie M. Faas, Johannes Kraus, Alexander Schoenhals, and Martin Baumann. 2021. Calibrating Pedestrians' Trust in Automated Vehicles: Does an Intent Display in an External HMI Support Trust Calibration and Safe Crossing Behavior?. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. 1–17.

D. Harrison Mcknight, Michelle Carter, Jason Bennett Thatcher, and Paul F. Clay. 2011. Trust in a Specific Technology: An Investigation of Its Components and Measures. *ACM Trans. Manage. Inf. Syst.* 2, 2, Article 12 (July 2011), 25 pages. https://doi.org/10.1145/1985347.1985353

Vivien Melcher, Stefan Rauh, Frederik Diederichs, Harald Widlroither, and Wilhelm Bauer. 2015. Take-over requests for automated driving. *Procedia Manufacturing* 3 (2015), 2867–2873.

Tim Miller. 2019. Explanation in artificial intelligence: Insights from the social sciences. *Artificial Intelligence* 267 (Feb. 2019), 1–38.

Sina Mohseni, Niloofar Zarei, and Eric D. Ragan. 2018. A Survey of Evaluation Methods and Measures for Interpretable Machine Learning. *CoRR* abs/1811.11839 (2018). arXiv:1811.11839 http://arxiv.org/abs/1811.11839

Nadia Mullen, Judith L. Charlton, Anna Devlin, and Michel Bédard. 2011. Simulator Validity: Behaviors observed on the simulator and on the road. In *Handbook of*

Driving Simulation for Engineering, Medicine, and Psychology, Donald L. Fisher, Matthew Rizzo, Jeffrey Caird, and John D. Lee (Eds.). CRC Press, Boca Raton, FL, USA, Chapter 13-1, 13–1–13–18.

Jakob Nielsen. 1994. Ten usability heuristics. (1994). Retrieved April, 2022 from https://www.nngroup.com/articles/ten-usability-heuristics

Brittany E Noah, Thomas M Gable, Shao-Yu Chen, Shruti Singh, and Bruce N Walker. 2017. Development and preliminary evaluation of reliability displays for automated lane keeping. In *Proceedings of the 9th international conference on automotive user interfaces and interactive vehicular applications*. 202–208.

Daniel Omeiza, Helena Webb, Marina Jirotka, and Lars Kunze. 2021. Explanations in Autonomous Driving: A Survey. *IEEE Transactions on Intelligent Transportation Systems* 23 (2021), 10142–10162. Issue 8.

P. Jonathon Phillips, Carina A. Hahn, Peter C. Fontana, David A. Broniatowski, and Mark A. Przybocki. 2020. *Four Principles of Explainable Artificial Intelligence.* Technical Report. https://doi.org/10.6028/NIST.IR.8312-draft

Andrea Papenmeier, Gwenn Englebienne, and Christin Seifert. 2019. How model accuracy and explanation fidelity influence user trust in AI. In *Proceedings of IJCAI 2019 Workshop on Explainable Artificial Intelligence (xAI)* (Macau, China). https://arxiv.org/pdf/1907.12652.pdf

So Yeon Park, Dylan James Moore, and David Sirkin. 2020. *What a Driver Wants: User Preferences in Semi-Autonomous Vehicle Decision-Making.* Association for Computing Machinery, New York, NY, USA, 1–13. https://doi.org/10.1145/3313831.3376644

Sebastiaan Petermeijer, Pavlo Bazilinskyy, Klaus Bengler, and Joost De Winter. 2017. Take-over again: Investigating multimodal and directional TORs to get the driver back into the loop. *Applied ergonomics* 62 (2017), 204–215.

Arianna Pipitone and Antonio Chella. 2021. What robots want? Hearing the inner voice of a robot. *iScience* 24 (2021). Issue 4. https://doi.org/10.1016/j.isci.2021.102371

Ioannis Politis, Stephen Brewster, and Frank Pollick. 2013. Evaluating multimodal driver displays of varying urgency. In *Proceedings of the 5th International Conference on Automotive User Interfaces and Interactive Vehicular Applications*. 92–99.

Ioannis Politis, Stephen Brewster, and Frank Pollick. 2015a. Language-based multimodal displays for the handover of control in autonomous cars. In *Proceedings of the 7th international conference on automotive user interfaces and interactive vehicular applications*. 3–10.

Ioannis Politis, Stephen Brewster, and Frank Pollick. 2015b. To beep or not to beep? Comparing abstract versus language-based multimodal driver displays. In *Proceedings of the 33rd annual ACM conference on human factors in computing systems*. 3971–3980.

Ioannis Politis, Stephen Brewster, and Frank Pollick. 2017. Using multimodal displays to signify critical handovers of control to distracted autonomous car drivers. *International Journal of Mobile Human Computer Interaction (IJMHCI)* 9, 3 (2017), 1–16.

Ioannis Politis, Stephen A Brewster, and Frank Pollick. 2014. Evaluating multimodal driver displays under varying situational urgency. In *Proceedings of the SIGCHI conference on Human Factors in Computing Systems*. 4067–4076.

Iyad Rahwan, Manuel Cebrian, Nick Obradovich, Josh Bongard, Jean-François Bonnefon, Cynthia Breazeal, Jacob W. Crandall, Nicholas A. Christakis, Iain D. Couzin, Matthew O. Jackson, Nicholas R. Jennings, Ece Kamar, Isabel M. Kloumann, Hugo Larochelle, David Lazer, Richard McElreath, Alan Mislove, David C. Parkes, Alex 'Sandy' Pentland, Margaret E. Roberts, Azim Shariff, Joshua B. Tenenbaum, and Michael Wellman. 2019. Machine behaviour. *Nature* 568 (2019), 477–486. Issue 7753. https://doi.org/10.1038/s41586-019-1138-y

Maria Riveiro and Serge Thill. 2021. "That's (not) the output I expected!" On the role of end user expectations in creating explanations of AI systems. *Artificial Intelligence* 298 (2021), 103507. https://doi.org/10.1016/j.artint.2021.103507

Alessandra Rossi, Kerstin Dautenhahn, Kheng Lee Koay, Michael L Walters, and Patrick Holthaus. 2020. Evaluating people's perceptions of trust in a robot in a repeated interactions study. In *International Conference on Social Robotics*. Springer, 453–465.

SAE On-Road Automated Vehicle Standards Committee and others. 2018. Taxonomy and Definitions for Terms Related to Driving Automation Systems for On-Road Motor Vehicles. *SAE International Warrendale, PA, USA* (2018).

Katri Salminen, Ahmed Farooq, Jussi Rantala, Veikko Surakka, and Roope Raisamo. 2019. Unimodal and multimodal signals to support control transitions in semiautonomous vehicles. In *Proceedings of the 11th International Conference on Automotive User Interfaces and Interactive Vehicular Applications*. 308–318.

Wojciech Samek and Klaus-Robert Müller. 2019. *Towards Explainable Artificial Intelligence.* Springer International Publishing, Cham, 5–22. https://doi.org/10.1007/978-3-030-28954-6_1

Tobias Schneider, Sabiha Ghellal, Steve Love, and Ansgar RS Gerlicher. 2021a. Increasing the User Experience in Autonomous Driving through different Feedback Modalities. In *26th International Conference on Intelligent User Interfaces*. 7–10.

Tobias Schneider, Joana Hois, Alischa Rosenstein, Sabiha Ghellal, Dimitra Theofanou-Fülbier, and Ansgar R.S. Gerlicher. 2021b. ExplAIn Yourself! Transparency for Positive UX in Autonomous Driving. 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 161, 12 pages. https://doi.org/10.1145/3411764.3446647

Tobias Schneider, Joana Hois, Alischa Rosenstein, Raimondo Lazzara, Steve Love, and Ansgar RS Gerlicher. 2021c. Velocity Styles for Autonomous Vehicles affecting Control, Safety, and User Experience. In *Symposium on Spatial User Interaction*. 1–2.

Martin Schrepp, Andreas Hinderks, and Jörg Thomaschewski. 2017. Design and Evaluation of a Short Version of the User Experience Questionnaire (UEQ-S). *Ijimai* 4, 6 (2017), 103–108.

Bobbie D Seppelt and John D Lee. 2007. Making adaptive cruise control (ACC) limits visible. *International journal of human-computer studies* 65, 3 (2007), 192–205.

Brian Stanton and Theodore Jensen. 2021. *Trust and Artificial Intelligence.* Technical Report. https://doi.org/10.6028/NIST.IR.8332-draft

Ariel Telpaz, Brian Rhindress, Ido Zelman, and Omer Tsimhoni. 2017. Using a vibrotactile seat for facilitating the handover of control during automated driving. *International Journal of Mobile Human Computer Interaction (IJMHCI)* 9, 3 (2017), 17–33.

Anouk van Maris, Hagen Lehmann, Lorenzo Natale, and Beata Grzyb. 2017. The influence of a robot's embodiment on trust: A longitudinal study. In *Proceedings of the Companion of the 2017 ACM/IEEE International Conference on Human-Robot Interaction*. 313–314.

Toni Wäfler and Ute Schmid. 2020. Explainability is not Enough: Requirements for Human-AI- Partnership in Complex Socio-Technical Systems. In *Proceedings of the 2nd European Conference on the Impact of Artificial Intelligence and Robotics (ECIAIR20)* (Portugal). Instituto Universitário de Lisboa (ISCTE-IUL), 185–193. https://doi.org/10.34190/EAIR.20.007

Marcel Walch, Kristin Lange, Martin Baumann, and Michael Weber. 2015. Autonomous driving: investigating the feasibility of car-driver handover assistance. In *Proceedings of the 7th international conference on automotive user interfaces and interactive vehicular applications*. 11–18.

Qianwen Wang, Yao Ming, Zhihua Jin, Qiaomu Shen, Dongyu Liu, Micah J. Smith, Kalyan Veeramachaneni, and Huamin Qu. 2019. ATMSeer: Increasing Transparency and Controllability in Automated Machine Learning. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (Glasgow, Scotland Uk) *(CHI '19)*. Association for Computing Machinery, New York, NY, USA, 1–12. https://doi.org/10.1145/3290605.3300911

Justin D. Weisz, Michael Muller, Stephanie Houde, John Richards, Steven I. Ross, Fernando Martinez, Mayank Agarwal, and Kartik Talamadupula. 2021. *Perfection Not Required? Human-AI Partnerships in Code Translation.* Association for Computing Machinery, New York, NY, USA, 402–412. https://doi.org/10.1145/3397481.3450656

Gesa Wiegand, Malin Eiband, Maximilian Haubelt, and Heinrich Hussmann. 2020. "I'd like an Explanation for That!"Exploring Reactions to Unexpected Autonomous Driving. In *22nd International Conference on Human-Computer Interaction with Mobile Devices and Services* (Oldenburg, Germany) *(MobileHCI '20)*. Association for Computing Machinery, New York, NY, USA, Article 36, 11 pages. https://doi.org/10.1145/3379503.3403554

Kathrin Zeeb, Axel Buchner, and Michael Schrauf. 2015. What determines the take-over time? An integrated model approach of driver take-over after automated driving. *Accident analysis & prevention* 78 (2015), 212–221.

A PROTOTYPE MATERIAL

Screenshots of online-questionnaire videos referencing to the situation numbers of Table 1.

(a) Driving view.

(b) Image shown when more information is requested.

Figure 6: Screenshot of the online-experiment for situation 1.

(a) Driving view.

(b) Image shown when more information is requested.

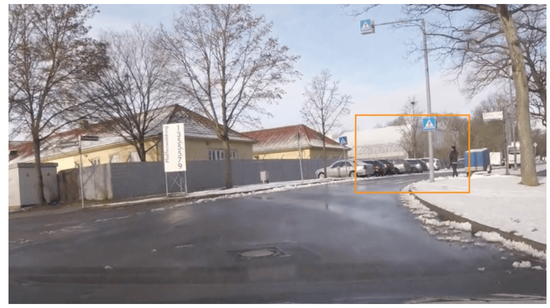
Figure 7: Screenshot of the online-experiment for situation 2.

(a) Driving view.

(b) Image shown when more information is requested.

Figure 8: Screenshot of the online-experiment for situation 3.

(a) Driving view.



(b) Image shown when more information is requested.

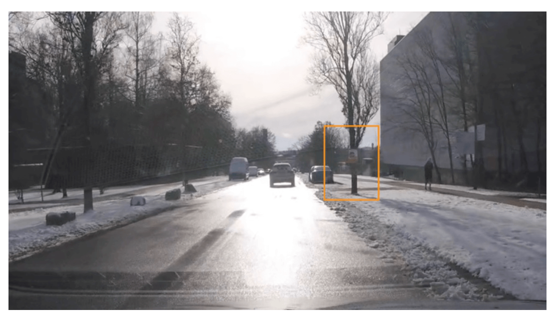
Figure 9: Screenshot of the online-experiment for situation 4.

(a) Driving view.

(b) Image shown when more information is requested.

Figure 10: Screenshot of the online-experiment for situation 5.

(a) Driving view.



(b) Image shown when more information is requested.

Figure 11: Screenshot of the online-experiment for situation 6.