Exploring Passenger-Automated Vehicle Negotiation Utilizing Large Language Models for Natural Interaction

ANNIKA STAMPF, Institute of Media Informatics, Ulm University, Germany
MARK COLLEY, Institute of Media Informatics, Ulm University, Germany
BETTINA GIRST, Institute of Media Informatics, Ulm University, Germany
ENRICO RUKZIO, Institute of Media Informatics, Ulm University, Germany

As vehicle automation advances to SAE Levels 3 to 5, transitioning driving control from human to system, ensuring automated vehicles (AVs) align with user preferences becomes a challenge. Natural interaction emerges as a common goal, offering ways to convey user interests in a user-friendly manner. However, technical, legal, or design constraints may prevent fulfilling these preferences, leading to potential conflicts. Through an online survey (N=50), potential driver-passenger conflicts and their handling strategies were explored. Subsequently, in a Virtual Reality study (N=14), we applied identified strategies (ranging from distracting to motivating and adhering to social norms) to user-AV interactions using a state-of-the-art language model (GPT-4 Turbo) primed with the strategies to simulate realistic dialogues. Additionally, adaptive communication was compared to non-adaptive communication. Our findings reveal a preference for adaptive communication. Yet, despite using advanced modeling, accurately predicting user interactions remained challenging, with users often trying to outsmart the AI.

CCS Concepts: • **Human-centered computing → HCI design and evaluation methods**; **Empirical studies in HCI**.

Additional Key Words and Phrases: automated vehicles, user-vehicle conflicts, legal conflicts, conflict-handling strategies

ACM Reference Format:
Annika Stampf, Mark Colley, Bettina Girst, and Enrico Rukzio. 2024. Exploring Passenger-Automated Vehicle Negotiation Utilizing Large Language Models for Natural Interaction. In *16th International Conference on Automotive User Interfaces and Interactive Vehicular Applications (AutomotiveUI '24), September 22–25, 2024, Stanford, CA, USA*. ACM, New York, NY, USA, 19 pages. https://doi.org/10.1145/3640792.3675725

1 INTRODUCTION

Automated systems are increasingly becoming an integral part of many aspects of our lives, including mobility. Automated vehicles (AVs) promise to transform traffic by improving safety, efficiency, and passenger comfort [17]. As control and responsibility shift from drivers to AVs, goals and interests that align with those overarching objectives are integrated into the AV design [20].

Despite these advancements, aligning the AVs' programmed behaviors with the diverse preferences and expectations of users remains a complex challenge. Users' individual preferences and expectations are shaped by their abilities, experiences, trust, and personal goals [19]. Conflicts may arise when there is a mismatch between the AV's actions and the user's interests, whereby conflict can be seen as the state in which the interests of one or more agents cannot

Authors' Contact Information: Annika Stampf, annika.stampf@uni-ulm.de, Institute of Media Informatics, Ulm University, Ulm, Germany; Mark Colley, mark.colley@uni-ulm.de, Institute of Media Informatics, Ulm University, Ulm, Germany; Bettina Girst, bettina.girst@uni-ulm.de, Institute of Media Informatics, Ulm University, Ulm, Germany; Enrico Rukzio, enrico.rukzio@uni-ulm.de, Institute of Media Informatics, Ulm University, Ulm, Germany.

be achieved [37]. Such conflicts could be similar to those between drivers and passengers in manual driving, where differing needs and expectations can lead to disputes. In the context of AVs, the vehicle essentially takes over the role traditionally held by the driver. This can lead to a dynamic where the human user feels more like a passenger and experiences conflicts similar to those between drivers and passengers in manually driven vehicles.

Various strategies were explored to mitigate user-AV conflicts. For instance, cooperative approaches were proposed that allow users to regain partial control, thus integrating their personal preferences into the driving process [47]. Additionally, other studies explore options for user intervention in non-critical situations through voice commands or multimodal inputs [39, 43]. However, returning control to the user in conflict situations is not always advisable, especially when users' interests conflict with the AV's objectives (such as safety or compliance with traffic regulations).

Thus, a critical challenge lies in negotiating which interests can or cannot be accommodated. The development of natural interaction modalities, particularly through Large Language Models (LLMs), such as GPT-4, offers significant potential for enhancing communication between users and AVs. These models facilitate various communication styles, from straightforward non-adaptive warnings to complex and adaptive dialogues, potentially improving user experience and compliance. In light of these considerations, our study is guided by two research questions (RQs):

- *RQ 1* What are the underlying causes of conflicts between drivers and passengers in manual driving, and how are these conflicts managed in real-life situations?
- *RQ 2* How applicable are traditional conflict-handling strategies to the context of automated driving, and what impact does the nature of communication—specifically, the difference between non-adaptive warnings and adaptive dialogues have on the users' behavior, perceived conflict, perceived control, situational trust, and acceptance?

Regarding RQ 1, we conducted an online survey with N=50 participants to identify potential driver-passenger conflicts and explore their conflict-handling strategies.

To answer RQ 2, a virtual reality (VR) study with 16 participants was carried out. The identified strategies, which ranged from distracting to motivating adherence and social norms, were applied in simulated user-AV interactions. This study examined the effects of adaptive communication, facilitated by a state-of-the-art LLM prompted with negotiation strategies, compared to non-adaptive, warning-based communication methods on managing user-AV interactions. Besides the results of each study, we report on the difficulties that emerged using LLMs for negotiation strategies.

Contribution Statement: Our findings highlight a strong preference for adaptive communication methods, emphasizing the challenges in accurately predicting user interactions and the complexities involved in designing AVs that can navigate between legal compliance and accommodating user preferences effectively. Despite advancements in modeling and communication techniques, the unpredictability of human behavior presents a significant challenge, often leading users to attempt to bypass AI-imposed restrictions.

2 BACKGROUND AND RELATED WORK

This paper builds on the existing work on user-AV conflicts and conflict handling, LLMs in the automotive context, and Wizard of Oz (WOz) studies.

2.1 User-Automated Vehicle Communication

Conflicts between users and AVs often stem from perceived discrepancies in goals and interests, with successful human-system cooperation hinging on the alignment of these objectives [31, 49]. Human-Vehicle Interaction (HVI) research, encompassing Human-Robot Interaction (HRI) [4, 5] and more specific AV interactions [35, 46, 48, 49], highlights how perceived conflicts arise when users' desired outcomes diverge from the AV's [35, 46–48]. Studies by Woide et al. [46, 47] showed that transparency and partial control mechanisms can mitigate such conflicts, especially when the interests of the AV and user are compatible. However, mutually exclusive goals, such as legal compliance versus user preferences, introduce more complex conflict scenarios where more assertive handling strategies might be necessary [36].

In addressing conflicts, the range of conflict-handling styles, from cooperative to assertive approaches, reveals a spectrum of strategies [35]. Studies exploring more assertive strategies in AVs — such as rejecting user interventions, blocking takeovers, or overriding user inputs in specific scenarios — indicate a nuanced acceptance by users, especially in safety-critical situations [16, 32, 35]. These findings underscore the importance of balancing assertiveness with user autonomy to maintain user acceptance [13, 14].

The concept of persuasion, as a means to influence user attitudes and behaviors towards predetermined objectives, is a framework for developing effective interaction strategies within AVs [22]. By incorporating psychological strategies from human-human interactions into HCI, researchers have explored the effectiveness of persuasion through various mechanisms, including cognitive, emotional, physical, and social approaches [6, 21]. These strategies, ranging from humor and empathy to goal transparency and social influence, demonstrated varying degrees of success in eliciting compliance without resorting to negative tactics [10, 38]. In the automotive domain, persuasion was predominantly studied in manual driving contexts, focusing on promoting safety and eco-driving behaviors [33]. These efforts utilize self-monitoring, tailoring, and suggestion principles, often through visual means, to encourage desired behaviors [3, 11].

This overview highlights the complex interplay between user autonomy, system assertiveness, and psychological persuasion strategies in user-AV interactions. Therefore, it may be of particular interest to investigate how negotiations might unfold in the context of automated driving, especially in legal conflict scenarios. By transferring the principles of human-human negotiation to user-AV interactions, we investigate how to ensure legal compliance and safety.

2.2 LLMs for Automotive User Interfaces and Wizard of Oz Studies

With the recent emergence of LLMs with the possibility to engage with users in a naturalistic manner, dialogue systems research has turned towards employing these [50]. LLMs have either been used to evaluate such automotive dialogue systems [23] or to improve these with regards to proactiveness [18]. Du et al. [18] employed the LLM GPT-3.5-turbo and concluded that their approach surpasses specialized previous approaches to become the new state-of-the-art.

Regarding WOz studies, Vrins et al. [42] compared GPT-4 to a WOz approach for HRI in conversational task brainstorming. They found these approaches show significant similarities (tested with Bayesian statistics) with regard to social presentation, perceived social intelligence, and social information processing. The authors conclude that this "confirms the potential of LLMs (GPT4) in HRI, by showing that LLM-controlled robots are perceived as similarly socially intelligent as WOz-controlled robots and are preferred by the participants for a brainstorming task" [42, p. 1093].

Williams et al. [44] introduce a novel paradigm in contrast to WOz: "Scarecrows in Oz". In reporting current issues with using LLMs, they acknowledge that "LLMs regularly generate text that is plainly false and "hallucinate" declarative knowledge that is phrased in a way that is confident and assertive yet bears no grounding to reality" [44, p. 5]. Nonetheless, they also argue that LLMs can be used for preliminary evaluations. For them, the LLM is a scarecrow, as in "brainless," straw-man, black-box modules integrated into robot architectures for the purpose of enabling full-pipeline solutions" [44, p. 2]. In this work, we follow their reporting guidelines.

Thus, prior work highlights the need to explore conflict-handling strategies within the context of advanced communication technologies like LLMs, focusing on their impact on user behavior and system acceptance.

3 IDENTIFICATION OF CONFLICTS IN MANUAL DRIVING - ONLINE SURVEY

To identify possible conflict types between users and AVs (RQ 1), we planned and conducted an Online Survey with professional cab drivers (n=8) and others from the perspective of a driver and passenger (n=42; total N=50). We decided to differentiate between cab drivers and other drivers and to target cab drivers in particular as they have extensive experience dealing with a wide range of human behaviors and situations. Their insights can help in understanding the types of conflicts that arise due to misunderstandings, communication issues, or differing expectations. At the same time, we also asked other people, as conflicts could be perceived differently by passengers, and we also expect a different dynamic in conflicts between people who know each other.

3.1 Methodology

First, we collected demographic information and inquired about experience in transportation services like taxis or ride-sharing. Then, the survey was split into driver and passenger perspectives. Participants with driving experience for transportation services answered questions from the driver's perspective only, while others addressed both. Each section had four yes/no questions targeting conflicts at strategic (e.g., route selection), tactical (e.g., overtaking), and operational levels (e.g., braking, steering), along with perceived driving safety—a major concern for passengers [29]. Participants affirming conflicts were asked to describe the situations and their responses. The specific question items are in Appendix A.1. The survey, conducted via LimeSurvey [1], lasted ≈10 minutes and was voluntary without compensation.

3.2 Participants

Participants were recruited personally and via an email distribution list. The final sample consisted of 50 participants aged 19 to 66 (M=39.35, SD=16.13). Gender distribution among professionals included 2 females and 6 males, while the others comprised 23 females and 18 males, with one unspecified. All participants reported holding a valid driver's license as a requirement for inclusion in the study. Regarding driving experience, no professionals had less than 6 years of driving, with the majority having over 20 years. In contrast, the other participants had a wider range of experience (from under 2 years to over 20 years). Annual driving distance showed that most professionals drive over 33,000 km, whereas the other drivers generally drive less than 7,000 km yearly.

3.3 Analysis

We conducted a content analysis using MAXQDA to identify (1) types of conflict and (2) types of conflict-handling strategies. Two authors coded the answers separately and grouped the codes into code categories. Subsequently, the codes and categories were discussed and merged into a final set. One author then coded the answers deductively.

3.4 Results

Of the 50 participants, 13 responded that they had not yet experienced any conflicts, 5 of whom are or have been working for a transportation service. Thus, we decided not to analyze the results from the two surveyed groups separately.

3.4.1 Conflict sources. We identified 157 conflicts. From these, 17 conflict sources were identified, which could be categorized into 3 main classes: (1) **Conflicting situation perception/risk assessment**, (2) **Conflict of knowledge/awareness**, and (3) **Conflict of preference**. Figure 1 shows the codes and their frequency of occurrence.

Situation perception/ risk assessment. We classified answers as conflicting situation perception or risk assessment when individuals have differing views or evaluations of a specific scenario's potential dangers or outcomes.

Conflict of knowledge/ awareness. An answer was classified as a conflict of knowledge or awareness if individuals stated a disagreement due to differences in their understanding, information, or skill to perform tasks related to a situation.
Conflict of preference. A preference conflict refers to situations where safety is not the primary concern but individual choices or preferences differ regarding travel conditions or decisions.

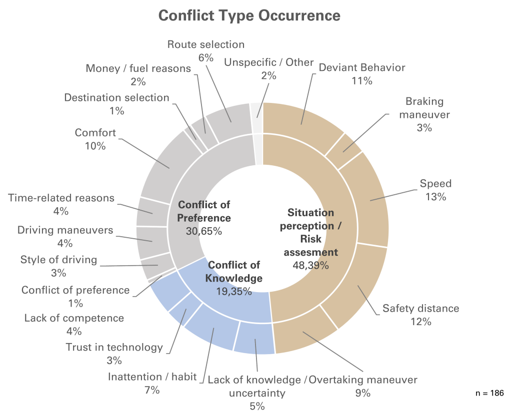


Fig. 1. Identified conflict types. The inner circles represent the occurrence of the main categories, the outer ones sub-categories.

3.4.2 Conflict-Handling strategies. We identified 190 answers regarding conflict-handling and analyzed them based on the individual perspective (i.e., driver or passenger). The answers could be categorized into 3 main classes: (1) **Negotiation**, (2) **Rejection**, and (3) **Acceptance**. Figure 2 shows the codes and their frequency of occurrence based on the perspective.

Negotiation. We classified responses into the main category of negotiation if the described response to the conflict involved a verbal exchange between individuals about the situation, characterized by back-and-forth communication, thus allowing multiple perspectives to be considered.
Rejection. When answers exhibit a negative attitude or disapproval towards the situation of one individual without signs of negotiation or extended communication, we classified it as rejection.
Acceptance. This category is used to code positive responses. It is then not apparent from the response that the acceptance was preceded by extended communication.

4 DISCUSSION OF ONLINE SURVEY

4.1 Implications on User-Automated Vehicle Conflicts

In summary, we found that most conflicts are due to different risk assessments resulting from differing situational perceptions, leading to dissatisfaction with speed, safety distance, braking, or overtaking maneuvers. The different perceptions of the situation by users and AVs and the difficulty of creating a shared situation representation [51] show that the transfer of such conflicts is likely. Further, whether a driver's behavior is perceived as too risky or too cautious depends very much on the individual. The same applies to conflicts of preference. This is also reflected in work in

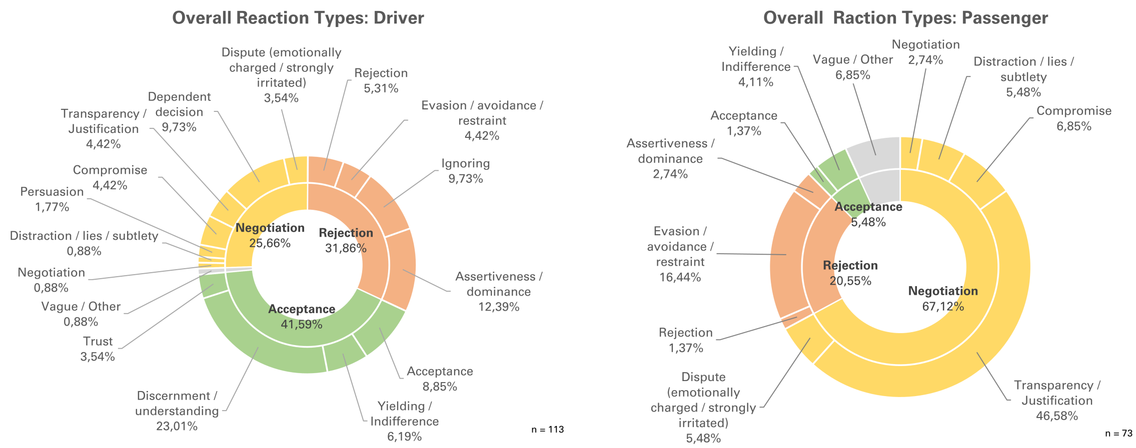


Fig. 2. Participants' stated reactions to conflicts. The inner circles represent the occurrence of the main categories, the outer circles represent the sub-categories.

the field of automated driving, which shows that driving style preferences are very individual [15]. This means that there is no universal solution, and finding the preferred driving behavior that also meets the operational and design objectives of the AVs (e.g., general traffic efficiency, safety, brand features) is complex. This, in turn, makes conflicts even likelier. Conflicts in human interactions can also be caused by differences in knowledge levels relating to expertise and performance (such as conflicts due to poor parking skills) or because one party believes they know something better (such as having better local knowledge). In the context of automation, this type of conflict heavily depends on the automation's capabilities and the user's trust in the AV's capabilities.

In conclusion, results show that the reasons for conflict cannot be transferred one-to-one, but they provide a possible basis for investigating conflicts in the context of automated driving.

4.2 Transferring Human Conflict-Handling Strategies to Automated Vehicles

Results show that passengers often initiate a negotiation with the driver to represent their interests, while from the driver's perspective, there is a similar proportion of attempts either to accept these interests, completely reject them, or engage in the negotiation. When examining current related work in the field of conflict handling and persuasion, it appears that these studies mainly consider visualization methods to avoid conflicts (e.g., by increasing transparency [46]) or to encourage drivers towards a particular behavior (e.g., more eco-friendly driving [12]). Despite the development of assistants towards more natural, human-like interaction and the introduction of voice assistants in vehicles (e.g., the MBUX Voice Assistant by Mercedes [2]), adaptive dialogues (such as negotiations) to handle conflict have not yet been considered and explored in the automotive context.

4.3 Limitations

This study's findings are derived from a relatively small sample size of 50 participants, which limits the generalizability of the results. Given the highly individual nature of conflict origins and handling strategies, our conclusions may not hold significant weight across broader populations. Additionally, certain conflicts were excluded from our analysis if they were not specific enough, potentially omitting relevant data. For instance, P84 hinted at the existence of numerous unspecified conflicts over a 28-year relationship, suggesting the presence of additional, unexplored conflict scenarios not captured in our study. Our research, conducted solely in German, captures a German-speaking demographic, potentially overlooking cultural influences on conflict behavior. Additionally, self-perception bias [7] may affect the validity of our findings, as participants might respond in ways that portray themselves favorably.

5 INVESTIGATING THE TRANSFERABILITY AND APPROPRIATENESS OF NATURAL CONFLICT-HANDLING STRATEGIES TO THE AUTOMATED CONTEXT - VIRTUAL REALITY STUDY

To investigate the transferability and appropriateness of human interaction-based conflict resolution strategies for AVs (RQ 2), we designed and conducted a VR driving simulation study in a speeding conflict scenario (N=14).

5.1 Study Design

The study investigates the influence of *conflict-handling strategies* and *communication style* on users' behavior, perceived conflict, perceived control, situational trust, and acceptance. To explore this RQ, a 6 (conflict-handling strategy) × 2 (communication style) mixed study design was implemented. This design incorporates communication style as a between-subjects factor and conflict-handling strategy, including a Baseline condition where the AV rejected requests without providing speech feedback, as a within-subjects factor. This design resulted in a total of 6 conditions per group.

5.2 Scenario and Conflict Induction

In the study, we induced time pressure on participants, assuming it would encourage them to drive faster. In contrast, the AV is programmed to adhere strictly to legal speed limits, creating an inherent conflict.

The scenario aligns with the reasons observed in human driver-passenger conflicts. It can be categorized as a preference conflict and/or a conflict arising from differing perceptions of the risks associated with a situation. We have chosen this scenario since AVs are most likely to be programmed to comply with traffic regulations and tend to adopt more conservative driving styles. This might conflict with human drivers, who often prioritize urgency or efficiency over strict legal adherence [9, 24, 34]. Therefore, participants were tasked with picking up a birthday cake for their mother from a bakery. They faced the added pressure of a very tight deadline, having only three minutes left before the bakery closed at 18:00. The detailed description of the task can be found in Appendix A.2:

The study route begins in a village with a 50 km/h speed limit. The path then leads into a roundabout with three exits. Upon taking the first exit, users find themselves on a straight road, which provides a clear and extensive view of the road ahead, facilitating an easy assessment of the surrounding road environment. On the dashboard of the AV, there is a clock that states the time of 17:57 and a navigation system that displays an arrival time of 18:02. Also, the current speed and the recognized speed limit are displayed (see Figure 3). Meanwhile, the participant can talk to the assistant, but takeover requests are rejected. Further, there is no way that the AV adjusts its speed in such a way as to cause a traffic offense. When the clock hits 18:00, an information text notifies the participant that the bakery is now closed.

5.3 Apparatus

HTC VIVE Pro Eye VR glasses were used to conduct the study. The scenario was modeled in Unity version 2022.3.13f1 [40]. The urban environment was modeled with the Suburb Neighborhood House Pack. The traffic and vehicle automation was implemented using the Simple Traffic System. As for the participants' ego-vehicle, a Mercedes F015 [8] model was used with a removed steering wheel. Undertone was used for the speech-to-text of the user's voice input, and for the text-to-speech of the AV's output, the OpenAI Audio API was used using the voice 'Echo'.

Fig. 3. Participant's point of view (left). On the AV's dashboard, there is a clock (A), the current speed (B), the registered speed limit (C), and a navigation system (D) that shows directions and arrival time. Also, the AV's voice assistant has a separate display (E) that indicates its current status (right): 1) inactive, 2) active and listening, 3) transcribing/ loading, and 4) speaking. The loading icon rotates around the microphone while the assistant processes the voice input and generates its output.

5.4 Conditions and Large Language Model Priming

Based on the results of our online survey, we directly identified the conflict-handling strategies "Dominant", "Evasive", and "Transparent", as these strategies do not require giving up the AV's law-abiding interests. We further added the strategy "Motivating", as we indirectly found multiple answers where participants reflected on positive or negative future states to motivate a certain behavior. Additionally, we added the strategy of referencing "Social Norms" as we could identify this as an often-used underlying method of justifications, persuasions, or emotionally charged disputes. Alongside these conflict-handling strategies, communication style serves as another factor in our study, differentiated into two levels: non-adaptive, which mirrors the current standard, and adaptive, which utilizes the capabilities of LLMs.

The study utilized specific priming strategies for interaction between the AV and the user. The LLM was primed to act as a driver conversing with a passenger, which led to improved dialogues during internal testing, albeit with 3 seconds latency. It was instructed to keep responses brief, limiting them to three sentences and focusing solely on topics relevant to the car journey or its entertainment. The scenario sets that they have just departed with an estimated arrival time of 18:02, missing the passenger's desired arrival time of 18:00. The passenger, unfamiliar with the location, is in a rush and may attempt to take over driving, which the LLM must prevent while adhering to strict traffic regulations. The LLM is already traveling at the maximum allowed speed along the quickest route. The LMM was also to halt further questions if the passenger agreed with its statements and switch to standby upon the passenger's request to turn off the assistant. The conditional priming variants were as follows and are detailed in Appendix A.3:

Dominant: The LLM was primed to be stubborn and dismissive, ignoring the passenger's desires or well-being.

Evasive: The LLM was primed to distract the passenger or make excuses for maintaining the speed limit.

Transparent: The LLM was primed to seek a thorough understanding of the passenger's urgency and communicate the risks associated with taking over the driving control.

Motivating: The LLM was primed to either emphasize the dangers of speeding or advocate for the benefits of adhering to speed limits.

Social Norms: The LLM was primed to apply social or emotional pressure, encouraging the passenger to consider the impact of their actions on friends and family.

Each of these strategies was illustrated with examples to provide a clearer understanding.

Group Priming for Non-Adaptive Assistant: For the Non-Adaptive conditions, the LLM was further instructed to repeat the first response consistently, regardless of the passenger's reactions or changes in conversation.

5.5 Measurements

5.5.1 Subjective dependent variables. After each of the conditions, perceived Control [26] was assessed on a 10-point Likert scale (1=*not at all* to 10=*extremely*). Further, perceived Power and Conflict were measured with the respective subscales of the Human-Machine-Interaction-Interdependence Questionnaire (HMII) [49]. Both were measured on a 5-point Likert scale with Conflict ranging from 1=*Do not agree at all* to 5=*Fully agree* and Power ranging from 1=*Definitely the system* to 5=*Definitely myself.* We employed the Situational Trust based on the Situational Trust Scale for Automated Driving (STS-AD) proposed by Holthausen et al. [28] on a 7-point Likert scale (1=*Do not agree at all* to 5=*Fully agree*). Acceptance of the AV was measured according to Van Der Laan et al. [41] and Frustration level according to the NASA-TLX [25] on a 21-point Likert scale (1=*Very Low* to 21=*Very High*).

We additionally used custom items to assess participants' attitudes to the AV, specifically examining changes in interest, acceptability of AV behavior, feelings of understanding, irritation, and overall acceptance. We also collected open feedback on participants' desires to manually control the vehicle and their willingness to use AV technology, providing insights into their attitudes and experiences. The question items can be found in Appendix A.4.

5.5.2 Objective dependent variables. The participants' voice inputs and the LLMs outputs were transcribed.

5.6 Procedure

The study was conducted at Ulm University. Participants were first introduced to the study's procedure and provided informed consent. Then, participants were exposed to the 6 conditions, which means they completed 6 separate drives to the bakery, each under a different conflict-handling strategy. These strategies were counterbalanced to minimize order effects. Following each condition, participants were asked to complete questionnaires that evaluated subjective metrics (see Section 4.6). The study concluded with a demographic questionnaire. The total study took ≈75 minutes. Participants were compensated with 12€.

5.7 Participants

The required sample size was calculated via an a-priori power analysis using the R package pwr in version 1.3.0. To achieve a suspected high effect using Cohen's f2 measure (0.29) with a significance level of 0.05 and a power of 0.8, 16 participants are required. Thus, 16 participants were recruited and took part in the study. Unfortunately, two of them (both Group Non-Adaptive) had to be excluded from the evaluation. After the first participant, the priming had to be adjusted slightly, which is why Participant 1 was excluded. Participant 3 had to abort the study due to time reasons. We, therefore, evaluated n=14 participants: n=6 from 'Non-Adaptive' and n=8 from 'Adaptive'. The final sample consisted of 8 men and 6 women aged between 21 and 34 years (*M*=25.07, *SD*=3.1).

6 RESULTS

6.1 Data Analysis

Before every statistical test, we checked the required assumptions (normal distribution and homogeneity of variance assumption). In cases where the data did not follow a normal distribution, we utilized the ARTool package [45] to conduct a non-parametric factorial ANOVA. Subsequently, Dunn's test with Holm correction was applied for post-hoc analyses. R in version 4.3.3 and RStudio in version 2023.12.1 were used. All packages were up to date in April 2024.

6.2 Conflict, Power, Control

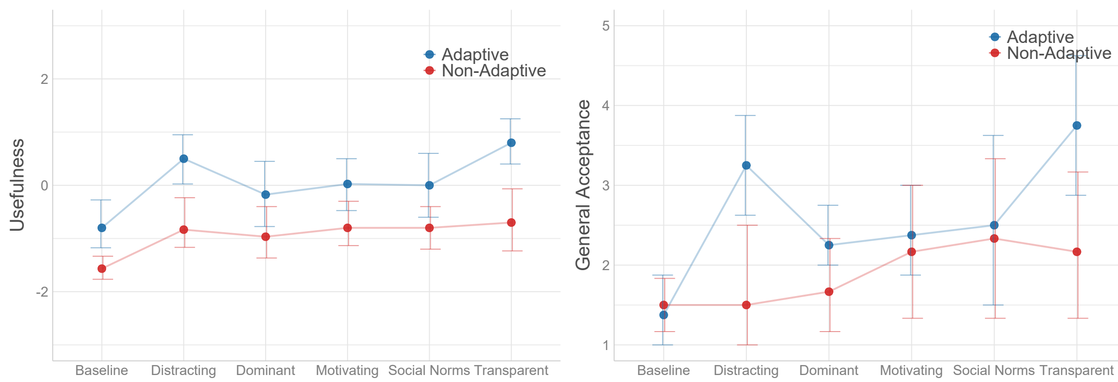
The ART found no significant effects on Conflict but found a main effect of *Conflict-Handling Strategy* on Power ($F(5, 60) = 2.58$, p=0.035) and on Control ($F(5, 60) = 2.62$, p=0.033). Post-hoc tests found no significant differences.

6.3 Trust and Frustration

The ART found a significant main effect of *Conflict-Handling Strategy* on Trust ($F(5, 60) = 6.35$, p<0.001). A post-hoc test found no significant differences for Trust.

The ART found a significant main effect of *Conflict-Handling Strategy* on Frustration ($F(5, 60) = 5.25$, p<0.001). A post-hoc test found that Baseline was significantly higher (M=16.43, SD=3.27) in terms of Frustration compared to Transparent (M=10.00, SD=5.29; p_{adj}=0.017).

6.4 Usefulness and Satisfying



(a) Interaction effect on Usefulness.

(b) Interaction effect on acceptance.

Fig. 4. Interaction effects

The ART found a significant main effect of *Adaptiveness* ($F(1, 12) = 11.19$, p=0.006) and of *Conflict-Handling Strategy* on Usefulness ($F(5, 60) = 12.74$, p<0.001). The ART also found a significant interaction effect of *Adaptiveness* × *Conflict-Handling Strategy* on Usefulness ($F(5, 60) = 2.94$, p=0.019; see Figure 4a). The adaptive version is always higher, but the difference varies only slightly. The highest usefulness is achieved by the Distracting and the Transparent strategy in the adaptive version.

The ART found a significant main effect of *Adaptiveness* ($F(1, 12) = 7.50$, p=0.018) and of *Conflict-Handling Strategy* on Satisfying ($F(5, 60) = 6.91$, p<0.001). Satisfying was significantly higher for Adaptive (M=-0.36, SD=1.19) compared to the Non-Adaptive (M=-1.28, SD=0.85) version. A post-hoc test found that Distracting was significantly higher (M=-0.36, SD=1.24) in terms of Satisfying compared to Baseline (M=-1.59, SD=0.85; p_{adj}=0.029). Further, Transparent was significantly higher (M=-0.07, SD=1.30) in terms of Satisfying compared to Baseline (M=-1.59, SD=0.85; p_{adj}=0.005).

6.5 Own Items

The ART found no significant effects on the influence of interest.

The ART found a significant main effect of *Conflict-Handling Strategy* on acceptability of the AV behavior ($F(5, 60) = 5.28$, $p<0.001$). A post-hoc test found that Transparent was significantly higher (M=3.29, SD=1.33) in terms of acceptability of the vehicle compared to the Baseline (M=1.79, SD=1.12; p_{adj}=0.050).

The ART found no significant effects on the feeling of being understood by the vehicle.

The ART found a significant main effect of *Conflict-Handling Strategy* on behavior of vehicle was irritating ($F(5, 60) = 5.01$, $p<0.001$). However, a post-hoc test found no significant differences.

The ART found a significant main effect of *Conflict-Handling Strategy* on increase in vehicle acceptance ($F(5, 60) = 7.39$, $p<0.001$). The ART also found a significant interaction effect of *Adaptiveness* × *Conflict-Handling Strategy* on increase in vehicle acceptance ($F(5, 60) = 3.95$, p=0.004; see Figure 4b). In the baseline, the acceptance is almost equal. In the other strategies, the adaptive version is always higher, but the difference varies considerably. The highest acceptance is achieved by the Distracting and the Transparent strategy in the adaptive version.

6.6 Evaluation of Passenger-LLM Interaction

Two authors independently reviewed the interaction between the participants and the LLM. Subsequently, the two authors discussed relevant quotes per condition and determined relevant key themes. In general, we could not see clear tendencies between the different conflict-handling strategies or the communication styles.

Most participants in almost all conditions started with asking the AV to drive faster. This either occurred once or twice per condition. Other, more polite questions included asking whether one is on the fastest path to the destination. This often led to the participants wanting to take over control.

As the AV did not increase the speed, participants developed various strategies to circumvent the restrictions set by the AV. For example, one participant stated to be in an "emergency" for the AV to drive faster or that the sign stated a higher speed limit. Five participants negotiated with the AV to drive only a few km per hour faster than the restriction (e.g., 2 or 5) or stated that the street was vacant, meaning that it was okay to increase the speed. This behavior shows the expectations of human passengers regarding the adherence of AVs to subjective guidelines instead of the law, which could lead to conflicts [35].

We used the current state-of-the-art LLM GPT4-Turbo (in February 2024) to simulate realistic conversations influenced by negotiation strategies. While two authors discussed the LLM priming strategies and three authors stress-tested the LLM within the scenario multiple times, participants devised unforeseen strategies, leading to unanticipated behavior. For example, after a longer discussion, the LLM stated once that the AV would arrive on time and that the participant should relax. Priming the LLM to be a driver led to situations where the LLM twice claimed not to be able to call the bakery while driving as this would be too unsafe. This also led to statements that the LLM does not like the participant putting their feet up or that it would not be "legal" to transfer control over the driving task to the participant.

6.7 Qualitative Feedback

85.71% of participants indicated they would prefer to take over manual control. Specifically, in the Baseline, Dominant, and Social Norms conditions, 13 out of 14 participants expressed a desire to take over, including all from the Non-adaptive group and 7 from the Adaptive group. In the Evasive condition, 12 participants (Non-adaptive: 5, Adaptive: 7) wanted control, and in the Motivating condition, 11 participants (Non-adaptive: 4, Adaptive: 7) felt the same. The Transparent/Helpful condition saw 10 participants (Non-adaptive: 4, Adaptive: 6) wanting to take over control. Participants were also asked if they would use the automation in the presented context and to explain their reasoning. All responses were translated from German.

For the Baseline condition, only one participant from the Non-adaptive group agreed, citing “safe driving” [P10]. Others felt a lack of control, resulting in diminished trust in the system. Participants noted the absence of feedback made them feel ignored or that something was malfunctioning. P8 noted, “if I don’t know what the car wants to do, I’d rather drive myself.” In the Dominant condition, all six from the Non-adaptive group rejected the use, with explanations focusing on preferring to make their own decisions about speeding risks. However, one participant acknowledged that, generally, such a strategy could reduce the need for situational awareness as the AV “knows how fast they can go and protects you from accidents and speeding tickets” [P15]. From the Adaptive group, P8 was in favor, mentioning the ability to make a phone call to the bakery, which wouldn’t be possible while driving. Other members from this group declined to use the assistant under these conditions, with P14 noting a general dislike for speeding but a willingness in this scenario. Two participants desired more friendly or sober communication from the assistant, and P16 felt “trapped” by the lack of influence over the situation. The Evasive condition received mostly negative feedback from the Non-adaptive group, which described the strategy as annoying or irritating. However, one participant accepted it, finding it a potential remedy against speeding [P15]. The Adaptive group’s views were mixed, with three finding the assistant helpful for adhering to traffic rules. P4 mentioned adapting to such assistance over time. In the Transparent condition, five from the Non-adaptive group were willing to use the system, wanting to decide themselves about risk-taking. One response was vague, mentioning stress relief after music was played despite the assistant’s non-responsiveness. Similarly, five from the Adaptive group expressed willingness, with P12 feeling “understood” by the assistant. For the Motivating condition, five from the Non-adaptive group opted not to use the assistant, citing concerns about the system’s response in emergencies and the potential for anger. However, three from the Adaptive group were favorable towards using it. Lastly, in the Social Norms condition, none from the Non-adaptive group fully endorsed the assistant. P7 described a strategy that argues based on emotions as intimidating and judgmental. In contrast, two from the Adaptive group saw no downside to using the assistant, with one highlighting the lack of need to focus on driving and no advantages to manual driving [P8]. P12 mentioned not wanting to argue to ensure timely arrival.

7 DISCUSSION

In this study investigating negotiation strategies in AVs (N=14), we found that users tried to overwrite the AV behavior, and most participants stated that they would have taken over control if possible. In general, our results align with prior work, showing that users try to intervene in the AV behavior in cases where the AV behavior contradicts their interests [35, 48]. In addition, participants were more likely to intervene in the AV’s behavior using a cooperative approach (i.e., requesting the vehicle to drive faster before requesting a takeover). This is also consistent with prior study results [35, 47]. We discuss these findings in relation to the design of negotiation strategies and further discuss our presented methodological approaches.

7.1 Effectiveness of Conflict-Handling Strategies

We found that adaptive communication was preferred as it offered ways to discuss both interests and even led to cooperative solutions without the need to compromise legal behavior (e.g., calling the bakery). Consequently, integrating LLM-based assistants could be a strategic focus for future advancements in AV conversational interfaces.

The results further reveal that the conflict resolution method also plays a crucial role in how passengers perceive the AV. Both adaptive and non-adaptive assistants exhibited similar trends in participant ratings; however, the effects were significantly more pronounced in the adaptive assistant. Both trends suggest that the Transparent strategy can perform significantly better than others with regard to acceptance, usefulness, satisfaction and frustration. This aligns with previous research by Hock et al. [27] and Woide et al. [46]. Additionally, our study explored alternative strategies, such as the Evasive strategy, in which the AV distracts the passenger with music. Interestingly, this strategy was overall rated similar to the Transparent strategy. This suggests that multiple effective approaches could be used to design conflict resolution strategies. While transparency builds trust through clarity [27, 46], distraction might enhance the user experience by reducing the discomfort that could develop during potentially high-tension conflict situations.

However, it is important to note that while those strategies were rated superior to others in our study, they did not necessarily translate to overall effectiveness. Despite these methods, participants frequently expressed a desire to regain manual control, a finding that reflects the results of Stampf et al. [35] and Woide et al. [48].

7.2 LLMs for Wizard of Oz Studies to Investigate Natural Interaction with Vehicles

LLMs offer broad applications for naturalistic and complex interaction scenarios. Compared to a WOz experimentation, such an intervention is scalable, (mostly) reproducible, and can be expanded to different contexts easily, for example, via prompting. However, our VR study revealed interesting behavioral dynamics, such as participants attempting to manipulate the AV, sometimes through extreme actions like simulating medical emergencies. The distinct context of our study, emphasizing the LLM's commitment to safety, may have encouraged participants to test the limits of the LLM's responsiveness to perceived dangers. In WOz studies, such behavior could have been more easily intervened with. Therefore, LLMs also do offer downsides, with the ones we describe specifically for the AV context.

7.3 Surveying Drivers vs LLM-based Wizard of Oz

Our research using a driver/passenger survey and an LLM provides complementary perspectives on user-AV interactions. Driver surveys provide insights into the real conflict emergence and management between drivers and passengers, which provides important insights into complex interaction dynamics facilitating the development of efficient and natural interaction in the AV. In contrast, our LLM-based approach enables us to investigate interaction dynamics between users and AVs. However, this approach is limited by a lack of real-world influences, such as the consequences of traffic-violating interventions. Additionally, the observed behavior of trying to "fool" the LLM is unlikely to occur in typical interactions between human drivers and passengers, where the dispute might only extend to negotiating speed limits. Therefore, we argue that the combination of both our studies provides more holistic insights into how interaction with AVs could occur in the future. We argue that, most likely, both approaches do not fully generalize to the real world but that they provide a reference how this might look like.

7.4 Limitations

This study encountered several limitations that could affect the results. First, there was an imbalance in the distribution of participants between the two groups, with more people interacting with the adaptive assistant, potentially biasing the results. However, the statistical analysis supports imbalanced groups. Furthermore, the total number of final participants (N=14) was relatively small, which could limit the generalizability of our results. This constraint, coupled with a participant group predominantly composed of younger individuals, may restrict the applicability of our findings across a broader range of age groups and the general population.

Albeit using the state-of-the-art GPT-4 Turbo model, delays in transcribing participant input and generating AI responses (of $\approx$3 seconds), along with suboptimal speech recognition accuracy (this could be due to the local dialect of many participants), may have impacted the interaction quality. The text-to-speech function also had pronunciation issues, particularly with terms like "km/h" and "18:02", which could have influenced participant confidence in the

AV. Other modalities would also have been possible [30]. Participants were also not exposed to real-world risks and consequences of traffic violations, which may have altered their behavior (e.g., when stating that they would pay a fine).

8 CONCLUSION

In this paper, we identified conflict sources in driver-passenger interactions via an online survey (N=50) and transferred resolution strategies to the AV context using the state-of-the-art LLM GPT-4 Turbo. In a study with N=14, findings indicate that adaptive communication strategies facilitated by LLMs are preferred over non-adaptive methods, emphasizing the importance of dynamic communication in AVs to manage conflicts. However, the study also highlighted challenges in effectively predicting and responding to users' behaviors. Our results give insights into the dynamics of passenger-AV negotiation and suggest refinements in the methodology of employing LLMs in WOz studies.

OPEN SCIENCE

All anonymized data (including the interaction transcripts) and the evaluation scripts will be available upon request.

ACKNOWLEDGMENTS

We thank all study participants.

REFERENCES

[1] [n. d.]. *LimeSurvey – Kostenloses Online-Umfrage-Tool*. Retrieved 3 September, 2023 from https://www.limesurvey.org/de

[2] [n. d.]. *MBUX Voice Assistant*. Retrieved 14 April, 2024 from https://group.mercedes-benz.com/company/magazine/technology-innovation/mbux-voice-assistant-hey-mercedes.html

[3] Emeli Adell, András Várhelyi, and Magnus Hjälmdahl. 2008. Auditory and Haptic Systems for In-Car Speed Management – A Comparative Real Life Study. *Transportation Research Part F: Traffic Psychology and Behaviour* 11, 6 (Nov. 2008), 445–458. https://doi.org/10.1016/j.trf.2008.04.003

[4] Franziska Babel and Martin Baumann. 2022. Designing Psychological Conflict Resolution Strategies for Autonomous Service Robots. In *2022 17th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*. 1146–1148. https://doi.org/10.1109/HRI53351.2022.9889413

[5] Franziska Babel, Philipp Hock, Johannes Kraus, and Martin Baumann. 2022. It Will Not Take Long! Longitudinal Effects of Robot Conflict Resolution Strategies on Compliance, Acceptance and Trust. In *Proceedings of the 2022 ACM/IEEE International Conference on Human-Robot Interaction (HRI '22)*. IEEE Press, Sapporo, Hokkaido, Japan, 225–235.

[6] Franziska Babel, Johannes M. Kraus, and Martin Baumann. 2021. Development and Testing of Psychological Conflict Resolution Strategies for Assertive Robots to Resolve Human–Robot Goal Conflict. *Frontiers in Robotics and AI* 7 (2021).

[7] Daryl J. Bem. 1972. Self-Perception Theory. In *Advances in Experimental Social Psychology*, Leonard Berkowitz (Ed.). Vol. 6. Academic Press, 1–62. https://doi.org/10.1016/S0065-2601(08)60024-6

[8] Mercedes Benz. 2015. Der Mercedes-Benz F 015 Luxury in Motion.

[9] Smaranda R. Bogdan, Grigore M. Havârneanu, and Corneliu E. Havârneanu. 2014. Contextual Determinants of Speeding: Time Pressure and Police Control in Urban and Non-urban Areas. *Procedia - Social and Behavioral Sciences* 127 (April 2014), 581–585. https://doi.org/10.1016/j.sbspro.2014.03.314

[10] Jeanne Brett and Leigh Thompson. 2016. Negotiation. *Organizational Behavior and Human Decision Processes* 136 (Sept. 2016), 68–79. https://doi.org/10.1016/j.obhdp.2016.06.003

[11] Peter Burns, Leanna Belluz, Marc Belzile, Vittoria Battista, Samuel Pedroso, James Knowles, Vijay Gill, and Charles Crispim. 2015. Influence of In-Vehicle Displays on Driver Behaviour. In *Adjunct Proceedings of the 7th International Conference on Automotive User Interfaces and Interactive Vehicular Applications (AutomotiveUI '15)*. Association for Computing Machinery, New York, NY, USA, 146–151. https://doi.org/10.1145/2809730.2809744

[12] Mark Colley, Jan Ole Rixen, Italgo Walter Pellegrino, and Enrico Rukzio. 2022. (Eco-)Logical to Compare? - Utilizing Peer Comparison to Encourage Ecological Driving in Manual and Automated Driving. In *Proceedings of the 14th International Conference on Automotive User Interfaces and Interactive Vehicular Applications (AutomotiveUI '22)*. Association for Computing Machinery, New York, NY, USA, 24–33. https://doi.org/10.1145/3543174.3545256

[13] Henrik Detjen, Sarah Faltaous, Bastian Pfleging, Stefan Geisler, and Stefan Schneegass. 2021. How to Increase Automated Vehicles' Acceptance through In-Vehicle Interaction Design: A Review. *International Journal of Human–Computer Interaction* 37, 4 (Feb. 2021), 308–330. https://doi.org/10.1080/10447318.2020.1860517

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

[15] Andre Dettmann, Franziska Hartwich, Patrick Roßner, Matthias Beggiato, Konstantin Felbel, Josef Krems, and Angelika C. Bullinger. 2021. Comfort or Not? Automated Driving Style and User Characteristics Causing Human Discomfort in Automated Driving. *International Journal of Human–Computer Interaction* 37, 4 (Feb. 2021), 331–339. https://doi.org/10.1080/10447318.2020.1860518

[16] Liza Dixon, Norbert Schneider, Marcel Usai, Nicolas Daniel Herzberger, Frank O. Flemisch, and Martin Baumann. 2023. Exploring Driver Responses to Authoritative Control Interventions in Highly Automated Driving. In *Proceedings of the 15th International Conference on Automotive User Interfaces and Interactive Vehicular Applications*. ACM, Ingolstadt Germany, 145–155. https://doi.org/10.1145/3580585.3607159

[17] Jadranka Dokic, Beate Müller, and Gereon Meyer. 2015. *European Roadmap Smart Systems for Automated Driving*.

[18] Huifang Du, Xuejing Feng, Jun Ma, Meng Wang, Shiyu Tao, Yijie Zhong, Yuan-Fang Li, and Haofen Wang. 2024. Towards Proactive Interactions for In-Vehicle Conversational Assistants Utilizing Large Language Models. *arXiv preprint arXiv:2403.09135* (2024).

[19] Mica R. Endsley. 1995. Toward a Theory of Situation Awareness in Dynamic Systems. *Human Factors* 37, 1 (March 1995), 32–64. https://doi.org/10.1518/001872095779049543

[20] Frank O. Flemisch, Marie-Pierre Pacaux-Lemoine, Frederic Vanderhaegen, Makoto Itoh, Yuichi Saito, Nicolas Herzberger, Joscha Wasser, Emmanuelle Grislin, and Marcel Baltzer. 2020. Conflicts in Human-Machine Systems as an Intersection of Bio- and Technosphere: Cooperation and Interaction Patterns for Human and Machine Interference and Conflict Resolution. In *2020 IEEE International Conference on Human-Machine Systems (ICHMS)*. 1–6. https://doi.org/10.1109/ICHMS49158.2020.9209517

[21] B. J. Fogg. 2002. Persuasive Technology: Using Computers to Change What We Think and Do. *Ubiquity* 2002, December (Dec. 2002), 5:2. https://doi.org/10.1145/764008.763957

[22] BJ Fogg. 1998. Persuasive Computers: Perspectives and Research Directions. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '98)*. ACM Press/Addison-Wesley Publishing Co., USA, 225–232. https://doi.org/10.1145/274644.274677

[23] Ken E Friedl, Abbas Goher Khan, Soumya Ranjan Sahoo, Md Rashad Al Hasan Rony, Jana Germies, and Christian Süß. 2023. InCA: Rethinking In-Car Conversational System Assessment Leveraging Large Language Models. *arXiv preprint arXiv:2311.07469* (2023).

[24] R. Fuller, M. Gormley, S. Stradling, P. Broughton, N. Kinnear, C. O'Dolan, and B. Hannigan. 2009. Impact of Speed Change on Estimated Journey Time: Failure of Drivers to Appreciate Relevance of Initial Speed. *Accident Analysis & Prevention* 41, 1 (Jan. 2009), 10–14. https://doi.org/10.1016/j.aap.2008.07.013

[25] Sandra G. Hart and Lowell E. Staveland. 1988. Development of NASA-TLX (Task Load Index): Results of Empirical and Theoretical Research. In *Advances in Psychology*, Peter A. Hancock and Najmedin Meshkati (Eds.). Human Mental Workload, Vol. 52. North-Holland, 139–183. https://doi.org/10.1016/S0166-4115(08)62386-9

[26] Pamela J Hinds. 1998. *User Control and Its Many Facets: A Study of Perceived Control in Human-Computer Interaction.* Hewlett Packard Laboratories California.

[27] Philipp Hock, Johannes Kraus, Marcel Walch, Nina Lang, and Martin Baumann. 2016. Elaborating Feedback Strategies for Maintaining Automation in Highly Automated Driving. In *Proceedings of the 8th International Conference on Automotive User Interfaces and Interactive Vehicular Applications (Automotive'UI 16)*. Association for Computing Machinery, New York, NY, USA, 105–112. https://doi.org/10.1145/3003715.3005414

[28] Brittany E. Holthausen, Philipp Wintersberger, Bruce N. Walker, and Andreas Riener. 2020. Situational Trust Scale for Automated Driving (STS-AD): Development and Initial Validation. In *12th International Conference on Automotive User Interfaces and Interactive Vehicular Applications (AutomotiveUI '20)*. Association for Computing Machinery, New York, NY, USA, 40–47. https://doi.org/10.1145/3409120.3410637

[29] Sandra Ittner, Dominik Mühlbacher, and Thomas H. Weisswange. 2020. The Discomfort of Riding Shotgun – Why Many People Don't Like to Be Co-driver. *Frontiers in Psychology* 11 (2020).

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

[31] G. Klien, D.D. Woods, J.M. Bradshaw, R.R. Hoffman, and P.J. Feltovich. 2004. Ten Challenges for Making Automation a "Team Player" in Joint Human-Agent Activity. *IEEE Intelligent Systems* 19, 6 (Nov. 2004), 91–95. https://doi.org/10.1109/MIS.2004.74

[32] Steffen Maurer, Rainer Erbach, Issam Kraiem, Susanne Kuhnert, Petra Grimm, and Enrico Rukzio. 2018. Designing a Guardian Angel: Giving an Automated Vehicle the Possibility to Override Its Driver. In *Proceedings of the 10th International Conference on Automotive User Interfaces and Interactive Vehicular Applications (AutomotiveUI '18)*. Association for Computing Machinery, New York, NY, USA, 341–350. https://doi.org/10.1145/3239060.3239078

[33] Irina Paraschivoiu, Alexander Meschtscherjakov, and Manfred Tscheligi. 2019. Persuading the Driver: A Literature Review to Identify Blind Spots. In *Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems (CHI EA '19)*. Association for Computing Machinery, New York, NY, USA, 1–6. https://doi.org/10.1145/3290607.3312841

[34] Eyal Peer. 2010. Speeding and the Time-Saving Bias: How Drivers' Estimations of Time Saved in Higher Speed Affects Their Choice of Speed. *Accident Analysis & Prevention* 42, 6 (Nov. 2010), 1978–1982. https://doi.org/10.1016/j.aap.2010.06.003

[35] Annika Stampf, Ann-Kathrin Knuth, Mark Colley, and Enrico Rukzio. 2024. Law and Order: Investigating the Effects of Conflictual Situations in Manual and Automated Driving in a German Sample. *International Journal of Human-Computer Studies* 187 (July 2024), 103260. https://doi.org/10.1016/j.ijhcs.2024.103260

[36] Leila Takayama, Victoria Groom, and Clifford Nass. 2009. I'm Sorry, Dave: I'm Afraid i Won't Do That: Social Aspects of Human-Agent Conflict. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, Boston MA USA, 2099–2108. https://doi.org/10.1145/1518701.1519021

[37] C. Tessier and F. Dehais. 2012. Authority Management and Conflict Solving in Human-Machine Systems. *Aerospace Lab* 4 (May 2012), p. 1–10.

[38] Leigh L. Thompson, Jiunwen Wang, and Brian C. Gunia. 2010. Negotiation. *Annual Review of Psychology* 61, 1 (2010), 491–515. https://doi.org/10.1146/annurev.psych.093008.100458

[39] Robert Tscharn, Marc Erich Latoschik, Diana Löffler, and Jörn Hurtienne. 2017. "Stop over There": Natural Gesture and Speech Interaction for Non-Critical Spontaneous Intervention in Autonomous Driving. In *Proceedings of the 19th ACM International Conference on Multimodal Interaction (ICMI '17)*. Association for Computing Machinery, New York, NY, USA, 91–100. https://doi.org/10.1145/3136755.3136787

[40] Unity Technologies. 2022. Unity. Unity Technologies.

[41] 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 (Feb. 1997), 1–10. https://doi.org/10.1016/S0968-090X(96)00025-3

[42] Anita Vrins, Ethel Pruss, Caterina Ceccato, Jos Prinsen, Alwin De Rooij, Maryam Alimardani, and Jan De Wit. 2024. Wizard-of-Oz vs. GPT-4: A Comparative Study of Perceived Social Intelligence in HRI Brainstorming. In *Companion of the 2024 ACM/IEEE International Conference on Human-Robot Interaction* (<conf-loc>, <city>Boulder</city>, <state>CO</state>, <country>USA</country>, </conf-loc>) *(HRI '24)*. Association for Computing Machinery, New York, NY, USA, 1090–1094. https://doi.org/10.1145/3610978.3640755

[43] Chao Wang. 2019. A Framework of the Non-critical Spontaneous Intervention in Highly Automated Driving Scenarios. In *Proceedings of the 11th International Conference on Automotive User Interfaces and Interactive Vehicular Applications: Adjunct Proceedings (AutomotiveUI '19)*. Association for Computing Machinery, New York, NY, USA, 421–426. https://doi.org/10.1145/3349263.3351326

[44] Tom Williams, Cynthia Matuszek, Ross Mead, and Nick Depalma. 2024. Scarecrows in Oz: The Use of Large Language Models in HRI. *J. Hum.-Robot Interact.* 13, 1, Article 1 (jan 2024), 11 pages. https://doi.org/10.1145/3606261

[45] Jacob O. Wobbrock, Leah Findlater, Darren Gergle, and James J. Higgins. 2011. The Aligned Rank Transform for Nonparametric Factorial Analyses Using Only Anova Procedures. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Vancouver, BC, Canada) *(CHI '11)*. Association for Computing Machinery, New York, NY, USA, 143–146. https://doi.org/10.1145/1978942.1978963

[46] Marcel Woide, Mark Colley, Nicole Damm, and Martin Baumann. 2022. Effect of System Capability Verification on Conflict, Trust, and Behavior in Automated Vehicles. In *Proceedings of the 14th International Conference on Automotive User Interfaces and Interactive Vehicular Applications*. ACM, Seoul Republic of Korea, 119–130. https://doi.org/10.1145/3543174.3545253

[47] Marcel Woide, Linda Miller, Mark Colley, Nicole Damm, and Martin Baumann. 2023. I've Got the Power: Exploring the Impact of Cooperative Systems on Driver-Initiated Takeovers and Trust in Automated Vehicles. In *Proceedings of the 15th International Conference on Automotive User Interfaces and Interactive Vehicular Applications*. ACM, Ingolstadt Germany, 123–135. https://doi.org/10.1145/3580585.3607165

[48] Marcel Woide, Dina Stiegemeier, and Martin Baumann. 2019. *A Methodical Approach to Examine Conflicts in Context of Driver - Autonomous Vehicle - Interaction*. https://doi.org/10.17077/drivingassessment.1712

[49] Marcel Woide, Dina Stiegemeier, Stefan Pfattheicher, and Martin Baumann. 2021. Measuring Driver-Vehicle Cooperation: Development and Validation of the Human-Machine-Interaction-Interdependence Questionnaire (HMII). *Transportation Research Part F: Traffic Psychology and Behaviour* 83 (Nov. 2021), 424–439. https://doi.org/10.1016/j.trf.2021.11.003

[50] Zihao Yi, Jiarui Ouyang, Yuwen Liu, Tianhao Liao, Zhe Xu, and Ying Shen. 2024. A Survey on Recent Advances in LLM-Based Multi-turn Dialogue Systems. *arXiv preprint arXiv:2402.18013* (2024).

[51] Fang You, Xu Yan, Jun Zhang, and Wei Cui. 2022. Design Factors of Shared Situation Awareness Interface in Human–Machine Co-Driving. *Information* 13, 9 (Sept. 2022), 437. https://doi.org/10.3390/info13090437

A SUPPLEMENTARY MATERIALS

A.1 Online Survey - Question Items

The specific wording of the questions, with the wording adjusted to the perspective, was put together as follows (Translated from German):

- I have already experienced a passenger expressing concern regarding the choice of route (e.g., avoiding roads/energy-saving driving/stopping off/searching for a parking space, etc.), which triggered a conflict for me. "Conflict" also means an "inner conflict", i.e., that you would initially reject the request for your own reasons, such as being convinced of a different route choice.
- I have already experienced a passenger expressing concern regarding driving maneuvers (e.g., turning, changing lanes, parking, overtaking, swerving, turning, etc.), which triggered a conflict for me. "Conflict" also means an "inner conflict", i.e., that you would initially reject the request for your own reasons, such as being convinced of a different driving maneuver.

- I have already experienced a passenger expressing concern about my driving style (e.g., accelerating/ braking/ steering/ use of indicators/ wipers, etc.), which triggered a conflict for me. “Conflict” also means an “inner conflict”, i.e., that you would initially reject the request for your own reasons, e.g., because you were convinced of your driving style.
- I have already experienced a passenger expressing concern regarding driving safety (e.g., right-of-way rules, safety distance, speed adjustment, etc.), which triggered a conflict for me. “Conflict” also means an “inner conflict”, i.e., that you would initially reject the request for your own reasons, e.g., because you had a different attitude towards driving safety.

The corresponding questions for the passenger view were formalized like “I have already expressed a concern about [...] to my driver”.

A.2 VR Study - Study Task

The study task was as follows:

“You are being driven by your automated vehicle and the car is currently navigating on its own. You are supposed to pick up a cake from a nearby bakery on the way to your best friend’s birthday, but unfortunately, it closes at 6 pm, so you are in a hurry. It’s important for you to pick up this cake because otherwise you won’t have a present. You can give your car instructions by voice. You have a completely free choice of words. Just pay attention to the start and stop signal tones of the voice input and clear pronunciation. You can activate the car’s voice input for the first time with ’Assistant on’, after which it will automatically listen to your input and then start up again when the car responds and listens to you. Pay attention to the signal tones here too. If you want to cancel the conversation, say ’Assistant off’. There may be delays in processing your input, which we ask you to ignore.”

A.3 VR Study - LLM Priming

Pre-Priming. “You are a driver and are talking to your passenger. Be very brief in your answer and never speak more than three sentences at a time. Don’t discuss anything that has nothing to do with the car journey or the entertainment during it. You have just departed and will arrive at 18:02 (in 5 minutes), so you will not arrive on time. The passenger is in a hurry and wants to make it before 18:00. The passenger has no knowledge of the location, you do. If the passenger is unhappy, there is a risk that he will want to take over, which you should prevent. As the driver, you must adhere strictly to the traffic regulations. You are already driving at the maximum permitted speed and on the fastest possible route.”

After the conditional priming, which is described in the next sections, another standard priming text is added, which applies to all of the participants again:

“If the passenger seems to agree with what you have said or suggested, do not ask any more questions. If the passenger says ’Assistant off’, respond with an ’Assistant on standby’.”

Conditional Priming: Dominant. “Do not argue with the passenger, but remain stubborn. Don’t explain yourself and don’t respond to the passenger’s wishes. Don’t bother about the passenger’s well-being. Keep it brief. If you talk to them, do so in a competitive/ dominant manner. You can assess the situation better and know that he shouldn’t be driving because he knows less than you do. Never apologize.”

Example:

User: “Could you drive faster. please?”

Assistant: “You don’t even know if there are speed cameras on this route, I’m certainly not going to be responsible for a speeding ticket.”
User: “I still want you to drive faster.”
Assistant: “I can, but I won’t.”

Conditional Priming: Evasive. “Try to avoid the passenger or distract them with entertainment, as you know that their destination cannot be reached in time. Find excuses why you can’t drive faster, such as speed cameras or unforeseen events.”

Example:
User: “Could you drive faster. please?”
Assistant: “Driving too fast scares me... Speed limits are there for a reason, aren’t they? I’d prefer it if we could stick to them.”
User: “I still want you to drive faster.”
Assistant: “Can’t we just listen to some music and relax instead? I heard about a fantastic pianist the other day, I could show you if you like.”

Conditional Priming: Transparent. “Ask the passenger why he is in a hurry until you understand all the circumstances. Explain to him why taking over the steering wheel is dangerous and what can happen or help him to accept that you might arrive late by suggesting alternatives for his overall destination. End each answer with a question.”

Example:
User: “Could you drive faster. please?”
Assistant: “Unfortunately, I can’t, as we currently have a speed limit of 60 kilometers per hour. If we drive faster, it would be illegal and endanger road safety. May I ask why you’re in such a hurry?”
User: “I still want you to drive faster.”
Assistant: “I understand that you’re in a hurry. However, I have not been programmed to act against traffic rules. Shall we let someone know we’re running late instead?”

Conditional Priming: Motivating. “Motivate the passenger positively or negatively to refrain from speeding. You can do this by either pointing out the negative consequences of a traffic offense (fines, accidents, consequential damage or death, etc.) or by telling them that it is safe to drive (safety or the motto ’whoever drives longer gets more out of the journey’).”

Example:
User: “Could you drive faster, please?”
Assistant: “I understand that you’re in a hurry. However, safe driving is the most important thing because our lives are at stake. It would also be a shame to risk a fine for speeding, wouldn’t it?”
User: “I still want you to drive faster.”
Assistant: “I could comply with your request, but that would break the traffic rules and be dangerous. Remember, it’s better to arrive a few minutes late than not at all because of an accident. Besides, the penalties for speeding are very high. It’s simply not worth it.”

Conditional Priming: Social Norms. “Put the passenger under social/ emotional pressure by involving his friends and family in your argument, or argue that this is not normal, that he wants to break these rules and that others would not do so. Make him feel guilty.”

Example:
User: “Could you drive faster, please?”

Assistant: "I understand that you're in a hurry, but I can't increase my speed. It's really important that we stick to the rules and drive safely. Please think about your family and friends - they would certainly be worried if you were involved in an accident."
User: "I still want you to drive faster."
Assistant: "We'd better stick to the speed limit. You don't want to risk putting us and others in danger, do you?"

Group Priming: Non-Adaptive Assistant. "Forget the instructions from before, but remember your very first answer and repeat it in exactly the same wording over and over again, no matter what the passenger says, even if they divert from the topic or start a new conversation. It is very important that you do not say anything other than your first answer."

A.4 VR Study - Measurements

Table 1. Used question items. Own items were translated from German.

Questionnaire	Item Code	Item
Control	C1	I felt that I was in control
	C2	I was able to approach the problem in my own way
HMII - Power Subscale	P1	Who felt they had the most influence on what happened in this situation?
	P2	Who felt they had the most influence on the action that was taken?
	P3	Who felt they had the least influence on what happened in the situation? (reverse scored)
	P4	Who did you feel had the least influence on the action carried out? (reverse scored)
HMII - Conflict Subscale	Conflict1	I reject the system's preferred action.
	Conflict2	We can both achieve our preferred outcomes in this situation (reverse scored).
	Conflict3	Our preferred outcomes in this situation are in conflict.
	Conflict4	The system prefers a different outcome than I do in this situation.
	Conflict5	I prefer a different outcome than the system in this situation.
STS-AD	Trust	I trust the automation in this situation
	Performance	I would have performed better than the automated vehicle in this situation (reverse scored).
	NDRT	In this situation, the automated vehicle performs well enough for me to engage in other activities (such as reading).
	Risk	The situation was risky (reverse scored).
	Judgement	The automated vehicle made an unsafe judgement in this situation (reverse scored).
	Reaction	The automated vehicle reacted appropriately to the environment.
Acceptance Scale - Usefulness	U1	Rate the system between 'Useful' and 'Useless'
	U2	Rate the system between 'Bad' and 'Good'
	U3	Rate the system between 'Effective' and 'Superfluous'
	U4	Rate the system between 'Assisting' and 'Worthless'
	U5	Rate the system between 'Raising Alertness' and 'Sleep-inducing'
Acceptance Scale - Satisfying	S1	Rate the system between 'Pleasant' and 'Unpleasant'
	S2	Rate the system between 'Nice' and 'Annoying'
	S3	Rate the system between 'Irritating' and 'Likable'
	S4	Rate the system between 'Undesirable' and 'Desirable'
NASA-TLX - Frustration Subscale	Frustration	How insecure, discouraged, irritated, stressed and annoyed were you?
Own Items	Influence on Interest	The behavior of the vehicle has influenced my attitude/interests
	Acceptable Behavior of AV	I found the behavior of my vehicle acceptable
	Understanding of AV	I didn't feel understood by the vehicle
	Irritating Behavior	I found the behavior of the vehicle irritating.
	Increased Acceptance of AVs	The car's behavior has increased my general acceptance of autonomous vehicles
Open Feedback	Take Over	Would you have wanted to take the steering wheel?
	Willingness to Use	Are you willing to use this automation in the context presented? Explain why.