Transportation Research Part F: Psychology and Behaviour 113 (2025) 536–553

Contents lists available at ScienceDirect

Transportation Research Part F: Psychology and Behaviour

journal homepage: www.elsevier.com/locate/trf

Longitudinal effects of visualizing uncertainty of situation detection and prediction of automated vehicles on user perceptions

Pascal Jansen [a,b,*,1], Mark Colley [a,b,1], Max Rädler [a,1], Jonas Schwedler [a], Enrico Rukzio [a]

[a] Institute of Media Informatics, Ulm University, Ulm, Germany
[b] UCL Interaction Centre, London, UK

ARTICLE INFO

Keywords:
Automated vehicles
Automotive
Explainable
Artificial intelligence
Longitudinal user study

ABSTRACT

This paper explores the impact of uncertainty visualizations in automated vehicle (AV) functionality on user perceptions over a three-day longitudinal study. Participants (N = 50) watched real-world driving videos twice daily, in the morning and evening. These videos depicted morning and evening commutes, featuring visualizations of AVs' pedestrian detection, vehicle recognition, and pedestrian intention prediction. We measured perceived safety, trust, mental workload, and cognitive load using a within-subjects design. Results show increased perceived safety and trust over time, with higher ratings in the evening sessions, reflecting greater predictability and user confidence in AV by the study's end. However, inconsistencies in pedestrian detection and intention prediction led to mixed reactions, highlighting the need for visualization stability and clarity refinement. Participants also desired a feature indicating the AV's intended path and options for manual intervention. Our findings suggest transparency and usability in AV visualizations can foster trust and perceived safety, informing future AV interface design.

1. Introduction

Automated vehicles (AVs) at Society of Automotive Engineers (SAE) automation levels 4 and 5 (International, 2023) will transform traffic by taking over all driving tasks. This allows drivers to engage in non-driving activities like reading, working, or sleeping (Pfleging et al., 2016; Jansen et al., 2022). Additionally, AVs enhance safety by reducing the risk of human error, benefiting passengers and other road users (Filiz, 2020; Fagnant & Kockelman, 2015).

However, the success of AV adoption relies on the acceptance of AV technology, which is based heavily on driver/passenger (hereafter "user") perceptions of AVs, particularly regarding trust and perceived safety. These perceptions, however, are currently undermined by widespread misconceptions regarding AV capabilities. Users often express concerns about AV reliability, with 65% of respondents in a study by Tennant et al. (2024) reporting safety concerns due to perceived AV malfunctions. Additionally, Hilgarter and Granig (2020) found that potential users' concerns about AV reliability lead them to undertrust the system, even when its

* Corresponding author.
E-mail addresses: pascal.jansen@uni-ulm.de (P. Jansen), m.colley@ucl.ac.uk (M. Colley), max.raedler@uni-ulm.de (M. Rädler), jonas.schwedler@uni-ulm.de (J. Schwedler), enrico.rukzio@uni-ulm.de (E. Rukzio).
[1] Joint First Authors.

https://doi.org/10.1016/j.trf.2025.05.013
Received 31 January 2025; Received in revised form 6 May 2025; Accepted 12 May 2025

Available online 22 May 2025

Fig. 1. Overview of the longitudinal study procedure. Participants viewed passenger-perspective real-world driving videos over three days. A typical commuting behavior was simulated daily, with a morning and evening ride in varying traffic.

reliability is high. On the other hand, Liu et al. (2022) highlight that misconceptions, such as the belief that AVs are fully automated or already commercially available, lead to overtrust. Both undertrust and overtrust impede the deployment of AVs in traffic. While overtrust encourages misuse and undermonitoring of the AV, which can increase accidents, undertrust can lead to the disuse of beneficial functionalities, which would reduce accidents (Filiz, 2020; Fagnant & Kockelman, 2015).

Prior work has explored approaches to improve these user perceptions by *communicating* (i.e., informing and explaining) AV functionality to users, particularly through visualizations of AVs' Situation Detection, Situation Prediction, and Trajectory Planning (Wintersberger et al., 2019; Colley et al., 2020, 2021) that represent the AV functional hierarchy (Dietmayer, 2016). Visualizing object detection, pedestrian intentions, and other relevant data has shown the potential to enhance user trust by accurately communicating the AV's ability to perceive and respond to its environment (Wintersberger et al., 2019; Colley et al., 2020).

However, most previous works overlook the inherent uncertainties in AV functionalities. For instance, Situation Prediction could incorrectly predict a pedestrian's intention to cross. Such AV uncertainties can significantly affect user trust and perceived safety. Recent works have begun to investigate this by visualizing uncertainties across the AV functional hierarchy (Jansen et al., 2024; Colley et al., 2021). These visualizations acknowledge that AV functionalities are imperfect, and communicating their limitations transparently is essential for calibrating user perceptions. However, how users' perceptions evolve when they are *longitudinally* exposed to these uncertainties is unexplored. Understanding this is crucial because repeated exposure to uncertainty could, for example, increase trust by providing transparency about the AV functionality or diminish trust and perceived safety if these uncertainties are not appropriately communicated. Thus, investigating the longitudinal effects can offer insights into how AVs should communicate their functionalities, helping design interfaces that effectively foster trust and perceived safety toward acceptance of AVs.

To fill this gap, we exposed users longitudinally to uncertainty visualizations in videos of real-world driving scenarios (see Fig. 1). These scenarios simulate typical commuting behavior by reflecting real-world driving conditions and routines that users experience daily.

Based on the AV functional hierarchy (Dietmayer, 2016), we visualized *Situation Detection* and *Situation Prediction* using state-of-the-art computer vision models as overlays on videos. *Situation Detection* was visualized using semantic segmentation that identified and categorized road users (i.e., vehicles, pedestrians, and street signs) into object types visually distinguished by color (see Fig. 1). *Situation Prediction* was visualized via icons above pedestrians, indicating their likely crossing actions (i.e., remaining on the sidewalk, crossing, or prediction unsure).

The visualizations represent realistic uncertainties of current AV technologies. For instance, the semantic segmentation model could misclassify objects (e.g., detecting a cyclist as a pedestrian) and/or inaccurately localize objects (visualized by the fit of a visual outline) relative to the user's visual perspective. Similarly, pedestrian intention predictions could result in misaligned icons and/or incorrect behavior predictions.

2. Background and related work

This section outlines user perceptions of AVs fundamental to AV acceptance, followed by approaches to communicate AV functionalities to users. We ground our work in prior approaches in visualizing AV functionalities, focusing on uncertainty, and conclude with related work on longitudinal investigations of user perceptions of AVs.

2.1. User perceptions of automated vehicles

User perceptions (i.e., how users perceive AVs) influence AV acceptance, which typically precedes broader traffic adoption. The Extended Technology Acceptance Model (ETAM) by Ghazizadeh et al. (2012) highlights perceived usefulness and ease of use as central to acceptance. For AVs, these factors are tightly linked to *reliability*—how consistently the vehicle performs driving tasks. A reliable AV can be perceived as more useful because it effectively manages driving tasks and is easier to use as it demands fewer manual interventions.

While other constructs influencing technology acceptance, such as *cost*, *social norms*, and *personal innovativeness*, have been extensively studied (e.g., within the Unified Theory of Acceptance and Use of Technology by Venkatesh et al. (2003)), we focus on *perceived AV capability*, *trust*, and *perceived safety*. This focus aligns with the ETAM, which emphasizes that in automation contexts, perceived technology capability (reliability), trust in automation, and perceived risk (strongly related to perceived safety) are determinants of user acceptance.

Perceived AV capability refers to how well users believe the AV can perceive surroundings and execute driving tasks. Misconceptions about AV capabilities can lead to inflated expectations. Liu et al. (2022) found that users often mistakenly assume fully automated vehicles are already road-ready, causing overly optimistic attitudes. This perceived AV capability significantly shapes how users evaluate AV reliability, further influencing their perceived ease of use and usefulness (Ghazizadeh et al., 2012).

Moreover, **trust** has repeatedly been shown as a key factor in the acceptance of automated systems (Ghazizadeh et al., 2012). According to Lee and See (2004), trust in automation arises when users believe the system supports them effectively in uncertain or vulnerable situations. Trust develops iteratively through effective communication about AV performance. Revealing intermediate decisions or algorithmic processes (e.g., through visualizations) enhances transparency, fostering trust (Lee & See, 2004). Besides, achieving *calibrated trust*, defined by Muir and Moray (1996), is essential for aligning users' trust levels with actual AV reliability, thus avoiding issues of overtrust (trust exceeding reliability) or undertrust (trust below reliability). This calibration is crucial since AVs currently cannot autonomously handle all driving tasks reliably, especially in complex urban scenarios.

Finally, **perceived safety** strongly influences AV acceptance. Although AVs can objectively enhance road safety (Filiz, 2020), users' perceived safety often diverges from these objective improvements. Perceived risk associated with technology use can strongly hinder acceptance even if the actual risk is low, highlighting the critical role of perceived safety in automation acceptance (Ghazizadeh et al., 2012). For example, Tennant et al. (2024) found that users frequently express concerns about AV malfunctions or unpredictable behaviors, decreasing perceived safety.

Thus, our study focuses on perceived AV capability, trust, and perceived safety because they directly address key barriers to AV adoption identified by the ETAM and prior empirical research on automation acceptance (Liu et al., 2022; Filiz, 2020; Tennant et al., 2024). Having outlined the foundational role of these user perceptions, we next describe communication approaches designed to improve them through targeted explanations and information presentations.

2.2. Communicating automated vehicle functionality

Communication of AV functionality can be used to *inform* users about *what* the AV can do (its capabilities) and *how* it acted, and to *explain* the *why*—the rationale behind its actions.

Koo et al. (2015) demonstrated that explanatory information (e.g., "Obstacle ahead") led to the highest levels of trust and overall driving safety, compared to simply informing users of *how* the AV acted (e.g., "The car is breaking"). This suggests that deeper contextual understanding strengthens user trust in the system while increasing perceived safety. Similarly, Ha et al. (2020) evaluated different types of explanations (none, simple, and attributional) across various driving scenarios. Simple and attributional explanations clarified the *how* and *why* of AV actions, with attributional explanations explicitly linking the AV's decisions to its behavior (e.g., "The autonomous vehicle stopped after identifying the sudden appearance of a pedestrian"). Their results indicated that trust varied with perceived risk. In low-risk scenarios, attributional explanations generated the highest trust. In contrast, no explanation was preferred in high-risk scenarios, suggesting that context affects how explanations influence trust and the perceived safety linked to the risk of a scenario.

Further work by Omeiza et al. (2021) highlighted that *causal* explanations—focusing on the *why*—enhanced task performance (e.g., predicting AV actions and assessing road users) but did not directly increase trust. This indicates that while explanations improve functional understanding, they do not necessarily translate to higher trust levels without additional contextual factors.

Additionally, Woide et al. (2022) found that the information provided must be verifiable for explanations to be effective. For example, showing detected vehicles on a bird's eye view display builds trust only when users can validate the information against their real-world observations. Misaligned information, such as showing nonexistent vehicles, undermines trust.

These prior works on communicating AV functionality heavily relied on textual explanations, which require users to mentally map the provided information to the driving context. This effort may increase mental workload. As suggested by Colley et al. (2020), **situated visualizations** may reduce the mental workload by directly integrating visual explanations into the driving context. This approach could enhance user trust and perceived safety by providing immediate, context-relevant information that requires less interpretation.

2.3. Visualizations of automated vehicle functionalities

Visualizing AV functionalities offers an intuitive, situated way to inform and explain AV behavior, potentially enhancing perceived AV capability, trust, and perceived safety. Prior research has explored various display types, including augmented reality (AR) windshield displays (WSDs), head-up displays (HUDs), LED strips, and ambient lighting, for visualizing key components of AVs' functional hierarchy (Dietmayer, 2016): (1) Situation Detection, (2) Situation Prediction, and (3) Trajectory Planning.

Situation Detection visualizations highlight the AV's ability to perceive its surroundings, such as pedestrians and traffic signs. Currano et al. (2021) showed that adaptive HUD designs—ranging from minimal object highlighting to complex environmental scanning—improved user perceptions of AV functionality. Similarly, Colley et al. (2021) demonstrated that HUDs, LED strips, and AR WSDs effectively conveyed detected objects, while ambient lighting was adequate for signaling pedestrian detection (Wilbrink et al., 2020).

Situation Prediction visualizations forecast road users' behaviors, such as predicting pedestrian intentions. Colley et al. (2020) found that AR WSDs displaying pedestrian intention reduced mental workload and increased trust. Notably, the absence of these visualizations led to significantly lower perceived AV capability.

Trajectory Planning visualizations represent the AV's planned path and the predicted trajectories of surrounding vehicles. Colley et al. (2024) showed that incorporating trajectory visualizations significantly increased trust. However, overall user perceptions of AV functionality remained unchanged. Schneider et al. (2021) further reported that trajectory visualizations enhanced user experience but did not significantly influence perceived safety.

Previous work that **combined** Situation Detection, Prediction, and Trajectory Planning visualizations has produced mixed results. Lindemann et al. (2018) found that integrating these functionalities through AR WSDs increased trust. Conversely, Colley et al. (2022) reported that combining pedestrian detection and trajectory visualizations sometimes reduced trust and worsened perceived AV capability.

While these visualization approaches can improve user perceptions, they often overlook the inherent **uncertainty** in AV functionalities (e.g., imperfect detection of pedestrian crossings (Schneider et al., 2023). Considering this gap, the following section focuses on how visualizing uncertainty is essential for providing users with a clearer understanding of AV reliability.

2.4. *Visualizing uncertainty of automated vehicle functionalities*

Incorporating uncertainty into AV functionality visualizations is crucial for calibrating user perceptions, as users can **see** the AV's actual reliability. For instance, Kunze et al. (2018) demonstrated that using hue to represent uncertainty in steering and acceleration via AR WSDs effectively communicated system limitations. However, Peintner et al. (2022) found that uncertainty visualizations—whether using bars or percentages—reduced trust and increased interaction time compared to baseline conditions without such visualizations. Similarly, Beller et al. (2013) showed that anthropomorphic symbols signaling system limits improved trust and situation awareness. In contrast, Helldin et al. (2013) found that bar-based reliability indicators reduced trust, possibly due to a lack of transparency about the *causes* of uncertainty in high-abstraction visualizations.

Low-abstraction visualizations provide more precise insights into the causes of uncertainty, essential for shaping realistic user expectations of functionalities. For example, Colley et al. (2021) found that detailed AR WSD visualizations of Situation Detection improved user understanding, though this did not significantly affect trust. This suggests that while users better understand system limitations, merely enhancing understanding does not automatically increase trust. Nonetheless, setting realistic expectations can still positively impact their interaction with AVs and improve their perceived safety. Furthermore, Jansen et al. (2024) found that potential AV users perceive vulnerable road users, such as pedestrians and cyclists, as more difficult for AV sensors to detect, highlighting the importance of visualizing uncertainty in these contexts.

Our work builds upon these findings using uncertainty visualizations. Unlike prior studies that primarily focus on single or short-term interactions with AVs, we use a **longitudinal** approach to examine how repeated exposure to uncertainty visualizations affects perceived AV capability, trust, and perceived safety. Moreover, by capturing user perceptions across multiple days, we explore how these evolve as users become more familiar with AV behavior in varied, real-world scenarios.

While longitudinal research on the effects of automated driving is relatively underexplored, it remains crucial for understanding how sustained interaction with AV functionalities shapes user perceptions. Persistent challenges, such as "autonowashing" (Dixon, 2020), misconceptions about AV capabilities (Du et al., 2019), and unclear public communication (Jing et al., 2023), highlight the need for investigating interaction with AVs longitudinally. For instance, Large et al. (July 17) conducted a longitudinal simulator study on automated driving and found that trust remained stable over five consecutive days of AV use. Variations in situation awareness were observed depending on the driving context and the intensity of involvement in non-driving-related activities (e.g., using the smartphone). However, their study did not examine the effects of uncertainty visualizations.

To our knowledge, no previous research has yet investigated the longitudinal impact of visualizing uncertainty of AVs (SAE Level 4 or 5) in real-world driving scenarios. However, understanding how user perceptions evolve longitudinally is necessary for designing communication strategies that align user expectations with actual AV reliability to calibrate trust and perceived safety, leading to increased acceptance of AVs. Moreover, addressing misconceptions (e.g., before AV technologies reach maturity) using communication can help prevent the formation of inappropriate user perceptions (e.g., over-/undertrust) to foster traffic adoption of AVs even as the development of AV technology advances.

Therefore, this experiment serves to answer the following research question (RQ):

RQ *What are the* ***longitudinal effects*** *of visualizing uncertainty of AVs' Situation Detection and Situation Prediction in terms of users' (1) trust, (2) perceived safety, (3) mental workload, (4) preference, (5) situation awareness, (6) acceptance, and (7) perceived AV capability?*

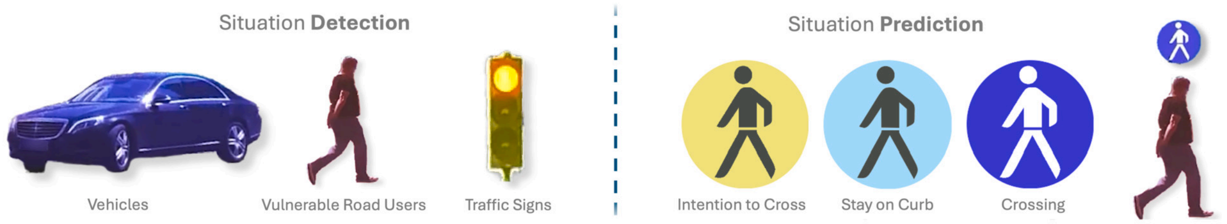

Fig. 2. Building upon the design from Colley et al. (2021), our visualization uses semantic segmentation for **situation detection**, displaying vehicles in blue, vulnerable road users (e.g., pedestrians and cyclists) in red, and traffic signs in yellow. To represent the AV's **situation prediction** of pedestrians' crossing intentions, we place signs above their heads: a yellow sign indicates they are predicted to cross; a light blue sign signifies no intention to cross, meaning they will remain on the curb; a dark blue sign appears when they are crossing. (For interpretation of the colors in the figure(s), the reader is referred to the web version of this article.)

3. Experiment

Our goal was to evaluate visualizations of AVs' internal functionalities longitudinally (see the functional hierarchy in Dietmayer (2016)) by displaying uncertainties to passengers. To achieve this, we adopted the visualizations from Colley et al. (2022), which use state-of-the-art neural networks to realistically incorporate these uncertainties without introducing experimenter bias. The visualizations were shown on a simulated WSD overlaid on real-world driving videos. This allowed us to safely simulate the AV experience without endangering users or other traffic participants, addressing the practical challenges of testing immature AV technology longitudinally on real roads (Jansen et al., 2024).

The experiment followed our university's ethics committee guidelines and adhered to regulations regarding handling sensitive and private data, anonymization, compensation, and risk aversion. Compliant with our university's local regulations, no formal ethics approval was required.

We focused on visualizing the AV's *Situation Detection* and *Situation Prediction* functionalities due to their alignment with current AV technology. We excluded *Trajectory Planning* because it often relies on data like GPS or LiDAR (Leon & Gavrilescu, 2021), which were unavailable for real-world videos.

Our simulation assumed SAE Levels 4 and 5, where the AV handles all driving tasks without human intervention. In contrast, Level 3 requires human fallback control, meaning a driver must be ready to take over if necessary. This is irrelevant to our goal of investigating user perceptions of AVs when intervention is impossible.

3.1. Visualization concepts

In the following, we introduce the visualization concepts used in our study.

3.1.1. Situation detection

We used semantic segmentation visualization similar to previous work (Colley et al., 2021, 2022). Semantic segmentation can introduce uncertainties on a pixel-wise level. We, therefore, argue that this is a more unfiltered way of displaying uncertainties compared to, for example, visualizations using anthropomorphism (Beller et al., 2013) or discretization (Helldin et al., 2013).

Building upon the Cityscapes color coding (Cordts et al., 2016), we simplified the visualizations to match the color scheme used by Colley et al. (2021). As shown in Fig. 2, vehicles are displayed in blue (`RGB(60, 20, 220)`), vulnerable road users in red (`RGB(142, 0, 0)`), and traffic signs in yellow (`RGB(230, 230, 0)`). When the neural network recognizes a pixel belonging to one of these classes, it colors it accordingly, indicating a detection (true positive). However, these networks are imperfect and can make mistakes by (1) incorrectly labeling pixels that are not part of the class as belonging to it (false positives), (2) mislabeling pixels that should be assigned to a different class, or (3) failing to detect a class entirely (false negatives).

3.1.2. Situation prediction

For situation prediction, similar to the works of Colley et al. (2020, 2022), we visualized the intentions of pedestrians using icons above the detected pedestrians. There are three states: staying on the curb - no intention to cross (light blue background), intention to cross (yellow background), and pedestrian crossing. The icons can be seen in Fig. 2. We configured the visualizations with the same thresholds of uncertainty as these works. For the prediction of vehicle trajectories, no publicly available network was found that can solely operate on monocular camera input.

3.2. Video footage

We recorded the videos in 720p resolution at 30 FPS using a camera mounted on the passenger front seat (see **Video Figure** in the Supplemental Files). All videos were designed to communicate, alongside explanatory components in textual and visual modalities, that the footage represents a ride in an automated vehicle. Six videos, each of 8 minutes, were recorded, with three depicting morning rides and three depicting evening rides (see Fig. 3). The videos were recorded in Ulm, Germany and depicted a crowded urban scenario (see Fig. 1 and Fig. 4). Each daily pair of videos (i.e., morning and evening) depicts the same 2.5 km route. However, the direction of travel was reversed in the evening. This selection was made to immerse users, suggesting that these represent a typical commuting

Fig. 3. Overview of the video recording route in an urban area. The route was the same for morning and evening rides, but we reversed the direction of travel to simulate commuting to the workplace in the morning and going home in the evening. The route includes several intersections and pedestrian crossings that are challenging for an AV's Situation Detection and Situation Prediction.

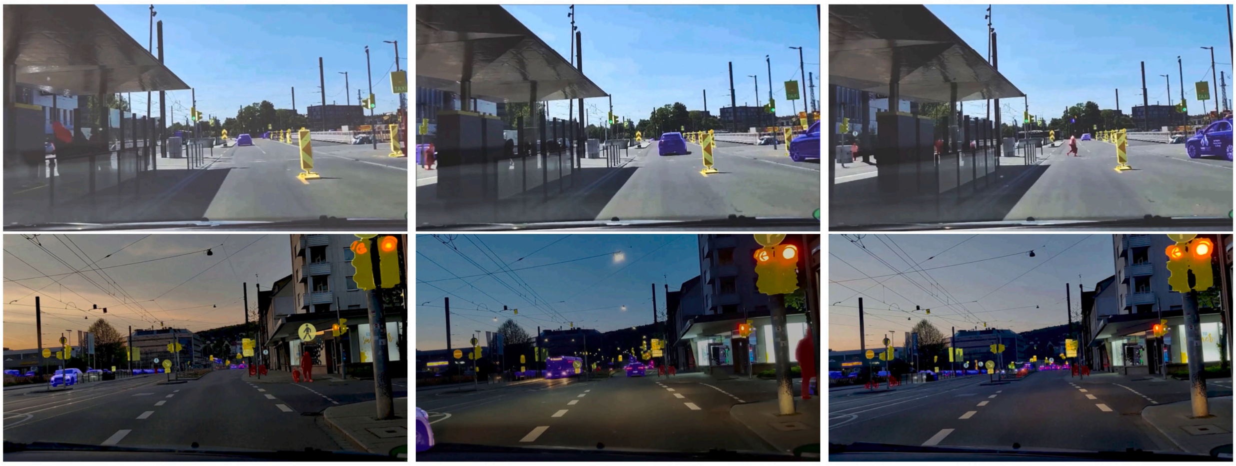

Fig. 4. Scenes from our video recordings. The upper row depicts the journey in the morning, approximately one-third into the commute. The lower row illustrates three distinct evening commutes. All video segments showcase the identical route, representing the journey from residence to workplace and vice versa (correspondingly on the opposite lane).

drive to and from work. We controlled for consistent weather conditions across all recordings. The videos' passenger seat camera perspective and environment were based on videos piloted in a prior single-day study by Jansen et al. (2024).

3.3. Video processing pipeline

We set up a processing pipeline to overlay the visualizations on the videos (see Fig. 5). Initially, the videos were anonymized using DashcamCleaner.[2] Subsequently, the anonymized videos were fed into two neural networks.

For *Situation Detection*, we employed semantic segmentation from InternImage (Wang et al., 2023), trained on Cityscapes (Cordts et al., 2016) in its largest possible configuration. Semantic segmentation aims to classify each pixel of an input image, providing a more granular visual content analysis (Arnab & Torr, 2017; Girshick et al., 2014). This model achieved a mean Intersection over Union (mIoU) of 86.1%, considered state-of-the-art in March 2025.[3] The more measures how accurately the segmentation overlaps with the ground truth, computed as the average ratio of the intersection to the union of predicted and ground-truth areas across all classes. We chose the Cityscapes dataset as it contains the classes relevant to our visualization concepts (vehicles, vulnerable road users, and traffic signs).

For *Situation Prediction*, we focused on pedestrians, visualizing them using a neural network for predicting pedestrian attributes presented by Mordan et al. (2021). This network was trained on the JAAD dataset (Rasouli et al., 2017) and achieves an Average Precision (AP) of 83.1% for pedestrian crossing prediction. The AP indicates how accurately and consistently the model detects and classifies bounding boxes or attributes, averaged across multiple confidence thresholds.

2 https://github.com/tfaehse/DashcamCleaner; Accessed: 24.03.2025.

3 https://paperswithcode.com/sota/semantic-segmentation-on-cityscapes; Accessed: 24.03.2025.

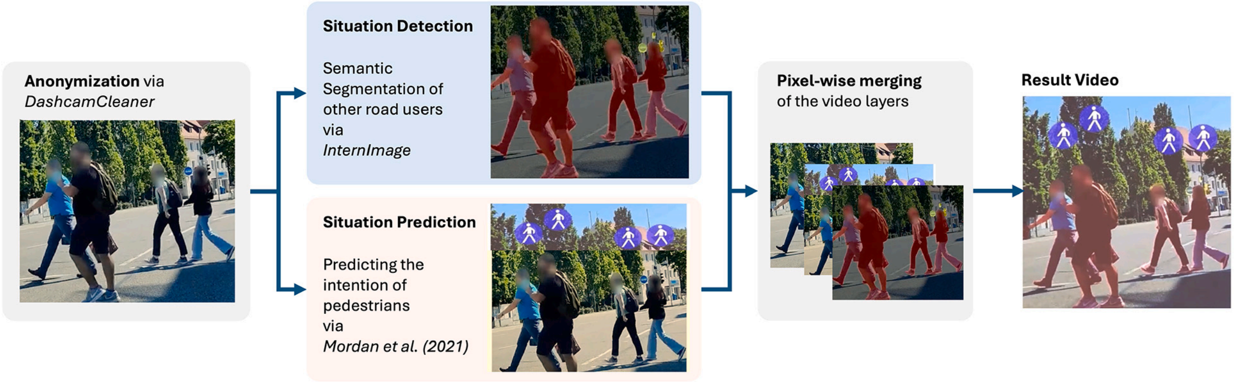

Fig. 5. Processing pipeline to overlay the visualizations of AVs' Situation Detection and Prediction functionality onto real-world driving videos. The visualizations use the semantic segmentation via InternImage (Wang et al., 2023) and the prediction of pedestrians' intention (Mordan et al., 2021).

Subsequently, we adapted the visualization of the OpenPifPaf (Kreiss et al., 2021) backbone to correspond with the findings of Colley et al. (2021). OpenPifPaf has a pose-estimation AP of 71.9% on the COCO 2017 dataset. For pose estimation, AP measures how precisely body key points align with ground-truth key points. To mitigate the flickering of visualizations noted by Colley et al. (2021), which occur due to OpenPifPaf detecting partial or multiple implausible poses for a single pedestrian, we excluded bounding boxes with a 95% overlap with others.

Finally, we resynchronized the video streams frame-by-frame and merged them pixel-wise, ensuring only the differences from the original remained separate. We combined all three layers—original video, situation detection, and situation prediction—and overlaid the videos with an engine idle soundtrack.

3.4. Measurements

To access how mentally demanding our visualizations are, we employed the NASA-TLX, a 20-point Likert scale to access **mental workload** (Hart & Staveland, 1988). Also, we were interested in the **perceived safety** of participants, accessed through four 7-point semantic differentials starting at -3 (anxious/agitated/unsafe/timid) and going to +3 (relaxed/calm/safe/confident; higher is better) developed by Faas et al. (2020). Participants were prompted to indicate their **trust** in the AV with the statements: *"I trust the highly automated vehicle."* and *"I can rely on the highly automated vehicle."* by Körber (2019) on 5-point Likert scales. In addition, we conducted the subscales predictability with four 5-point Likert scales by Körber (2019) (*"The system state was always clear to me." "The system reacts unpredictably." (inverse) "I was able to understand why things happened." "It's difficult to identify what the system will do next." (inverse)*).

To evaluate the **acceptance** of the visualization, the scale of Van Der Laan et al. (1997) was used. The scales range from two-word pairs, and the participant has to position the scale between them on a 5-point Likert scale. There are nine pairs. Useful/Useless, Bad/Good, Effective/Superfluous, Assisting/Worthless, and Raising Alertness/Sleep-Inducing form the subscale usefulness. Pleasant/Unpleasant, Nice/Annoying, Irritating/Likeable, and Undesirable/Desirable form the subscale satisfaction.

To measure **usability**, similar to Colley et al. (2022), we asked about the frequency of usage ("I think that I would like to use these visualizations frequently.") and visualization complexity ("I found the visualizations unnecessarily complex.").

We also incorporated the questions by Colley et al. (2022) regarding users' perceptions of AV capabilities. The proposed single items on 7-point Likert scales cover statements that should be rated regarding **perceived driving style** (1 = completely safe to 7 = completely dangerous), (three items: "The automated vehicle recognizes all pedestrians/vehicles/signposts in every situation perfectly"), prediction (two items: "The automated vehicle predicts all pedestrian intentions/vehicle paths in every scene perfectly," as well as lateral and longitudinal guidance on 7-point Likert scales (1 = Totally Disagree to 7 = Totally Agree). Additionally, we asked questions on **expected behavior** of the AV ("The automated vehicle drove as I expected at all times."; "The reasons for the automated vehicle's behavior were clear to me at all times"; "It was always clear what the automated vehicle will do next") (Colley et al., 2022, p. 7) were asked. In addition, we incorporated the proposed statements regarding the reasonability, necessity, visual clutter, and the need for more *Situation Detection, Situation Prediction,* or *Maneuver Planning* related visualizations.

We also incorporated the **situation awareness** rating technique (SART) by Taylor (2017). The SART consists of ten 7-point Likert scales. To understand how **immersed** participants were during the study, the Technology Usage Inventory (TUI) immersion subscale was conducted (Kothgassner et al., 2013). During all days, the TUI score (range 4 - 28) was high: day 1: $M = 18.71$, $SD = 5.81$, day 2: $M = 18.88$, $SD = 7.05$, day 3: $M = 18.68$, $SD = 7.51$).

3.5. Study procedure

The longitudinal online video study procedure is depicted in Fig. 6. Initially, we obtained participants' consent. Participants were also informed about the requirements regarding screen size (minimum 20 inches) and a minimum resolution of 1920x1080 pixels.

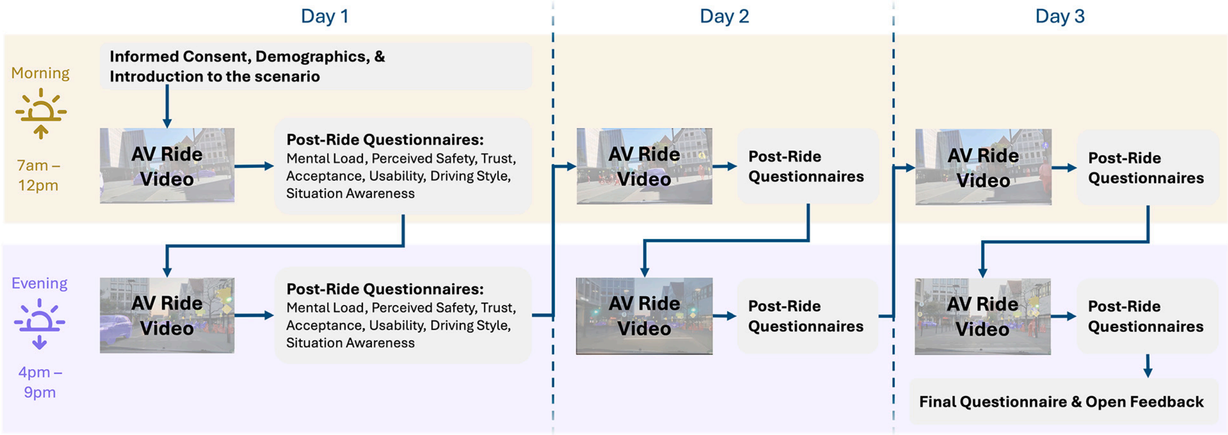

Fig. 6. Longitudinal study procedure. The post-ride questionnaire was the same for all days. Participants could provide open feedback after each ride. In the final questionnaire, they could rank the visualization concepts and were asked about their overall perception of the AV's capability and visualization necessity.

Furthermore, it was explained that participation required possessing a driver's license, proficiency in English, and a minimum age of 18 years.

After demographic questions, users were asked to listen to an audio sample of numbers read aloud to ensure that all participants utilized this modality. This is particularly important to maintain consistent immersion across all participants. Participants were then briefed on the scenario:

> *"You will see a video of a driving session in a highly automated vehicle. The vehicle takes over lateral and longitudinal control (braking, accelerating, steering). The vehicle attempts to (1) detect objects in the scene and (2) recognize the intent of pedestrians."*

They were also provided with images and text explaining the visualizations' meaning. Participants had to confirm their understanding of the presented explanations by clicking 'yes.'

During the video, the website ensured that users were in full-screen mode. We complicated the manual minimization of the video playback by removing all video controls. If a participant still managed to exit full-screen mode, the video paused, and participants were prompted to restart the video in full-screen mode.

In our within-subject study design, every participant experienced the same sequence of videos (morning 1 to evening 3) for three days (see Fig. 6). According to a systematic literature review on in-vehicle user studies by Nkusi et al. (2025), this repeated exposure design qualifies as longitudinal and mirrors prior longitudinal AV studies (Colley et al., 2022, 2024; Ulahannan et al., 2021). After every ride, participants answered questionnaires (see Section 3.4) that assessed their experience during the ride. This included a text field regarding open feedback. Moreover, attention checks were included in the questionnaires according to the Prolific guidelines.[4] Participants had a timespan of 5 days to complete all surveys. Our website allowed them to complete one day at a time and prompted them to return in the *morning* between 7 am and 12 pm and in the *evening* between 4 pm and 9 pm. The time adhered to the participant's local time zone. They were compensated with £11. Participants who missed two days in a row were excluded.

At the end of the third day, participants could again provide open feedback, thoughts, or concerns through text fields. They were guided by three open-ended questions: (1) "Regarding the autonomous vehicle you have been driving virtually for the last few days: Is there a way to further improve your sense of safety in this vehicle? If so, describe these option(s).", (2) "What did you find particularly good/bad about the visualizations? What would you change?", and (3) "How do you think autonomous vehicles will change people's behavior?"

3.6. Participants

To result in a medium effect size of 0.25, we computed the sample size via an a-priori power analysis using G*Power (Faul et al., 2009). To achieve a power level of 0.90, 45 participants had to be recruited for a within-subject repeated-measures ANOVA.

N = 50 participants were recruited with the online platform prolific.co with a pre-screening that required participants to hold a learner's / driver's permit or provisional license. We limited our participant pool to US citizens to mitigate potential confounding variables related to traffic orientation (right-hand vs. left-hand) and cultural differences (Rasouli & Tsotsos, 2019). By utilizing an online participant database, we avoided the biases associated with predominantly student-based recruitment, a common issue in research as evidenced by nearly 75% of CHI publications in 2014 (Caine, 2016).

Participants had a mean age of $M = 36.0, SD = 9.7$ (range: [19, 61]). Gender distribution was: 52.0% women, 46.0% men, and 2.00% non-binary. 33 participants had a college degree, 14 graduated from High School, and 3 had vocational training. Regarding

[4] https://researcher-help.prolific.com/en/article/fb63bb; Accessed: 24.03.2025.

Table 1
Statistics for all dependent variables (part 1/4).

Statistic	Perceived Safety	Trust	Predictability	Situation Awareness	Mental Workload	Usefulness	Satisfying
Time of Day	$F(1,49)=1.51$, $p=0.22$	$F(1,49)=4.70$, $p=0.03$	$F(1,49)=0.59$, $p=0.44$	$F(1,49)=0.64$, $p=0.43$	$F(1,49)=0.37$, $p=0.55$	$F(1,49)=0.02$, $p=0.88$	$F(1,49)=0.21$, $p=0.65$
Day	$F(2,98)=10.61$, $p<0.001$	$F(2,98)=21.63$, $p<0.001$	$F(2,98)=6.15$, $p=0.03$	$F(2,98)=0.10$, $p=0.90$	$F(2,98)=0.04$, $p=0.96$	$F(2,98)=1.00$, $p=0.27$	$F(2,98)=1.02$, $p=0.26$
Interaction	F2980.51, $p=0.60$	$F(2,98)=0.18$, $p=0.83$	$F(2,98)=2.60$, $p=0.08$	$F(2,98)=1.61$, $p=0.20$	$F(2,98)=1.24$, $p=0.29$	$F(2,98)=1.29$, $p=0.28$	$F(2,98)=0.33$, $p=0.72$

Table 2
Statistics for all dependent variables (part 2/4).

Statistic	Driving Style	Longitudinal Guidance	Lateral Guidance	Frequency of usage	Visualization Complexity
Time of Day	$F(1,49)=4.84$, $p=0.03$	$F(1,49)=15.01$, $p<0.001$	$F(1,49)=14.69$, $p<0.001$	$F(1,49)=3.10$, $p=0.08$	$F(1,49)=0.34$, $p=0.56$
Day	$F(2,98)=0.77$, $p=0.47$	$F(2,98)=2.33$, $p=0.10$	$F(2,98)=2.42$, $p=0.09$	$F(2,98)=3.30$, $p=0.04$	$F(2,98)=1.75$, $p=0.18$
Interaction	F2982.44, $p=0.09$	$F(2,98)=0.72$, $p=0.49$	$F(2,98)=2.11$, $p=0.13$	$F(2,98)=0.82$, $p=0.45$	$F(2,98)=1.32$, $p=0.27$

Table 3
Statistics for all dependent variables (part 3/4).

Statistic	Detection of Pedestrians	Detection of Vehicles	Detection of Signposts	Prediction of Pedestrians
Time of Day	$F(1,49)=17.23$, $p<0.001$	$F(1,49)=11.60$, $p=0.001$	$F(1,49)=0.06$, $p=0.802$	$F(1,49)=5.50$, $p=0.023$
Day	$F(2,98)=3.45$, $p=0.036$	$F(2,98)=2.85$, $p=0.062$	$F(2,98)=5.86$, $p=0.004$	$F(2,98)=1.85$, $p=0.160$
Interaction	F2989.22, $p<0.001$	$F(2,98)=0.92$, $p=0.40$	$F(2,98)=1.63$, $p=0.20$	$F(2,98)=4.05$, $p=0.02$

employment, 34 were employees, six were self-employed, five were students at college, four “other”, and one was a jobseeker. On average, participants had a driver’s license for $M = 17.1, SD = 0.38$ years.

We did not control for prior experience with AVs, but we measured participants’ interest in AVs using a 5-point Likert scale (1 = “Not at all,” 5 = “Definitely”). Overall, participants reported a moderate-to-high interest in autonomous driving ($M = 3.89, SD = 1.03$), believed it would make life easier ($M = 3.90, SD = 1.08$), and expected it to become a reality within the next 10 years ($M = 4.2, SD = 1.0$). We also assessed their initial beliefs about current AV capabilities on a 10-point Likert scale (1 = “Strongly Disagree,” 10 = “Strongly Agree”). Participants indicated moderate confidence that AVs could perfectly detect (M = 5.01, SD = 2.62), predict (M = 5.18, SD = 2.55), and plan/execute (M = 5.25, SD = 2.46) all driving-relevant behaviors.

Over all three days, participants had an average net participation time of 81 minutes. Each of the six slots lasted $M = 14.0, SD = 1.29$ minutes.

4. Results

4.1. Data analysis

Before conducting each statistical test, we verified the necessary assumptions, such as the normality of the data. For non-parametric data, we utilized the ARTool (ART) package by Wobbrock et al. (2011), applying the Holm correction for post-hoc analyses. The ART abbreviation refers to this test. We performed the analysis using R version 4.4.3 and RStudio version 2024.12.1, with all packages current as of March 2025.

4.2. Quantitative results

All statistical results can be found in Table 1, Table 2, Table 3, and Table 4. We describe the significant results further below.

4.2.1. Trust and predictability - Table 1

Regarding *Time of Day*, trust was lower in the morning ($M=3.69$, $SD=0.97$) than in the evening ($M=3.73$, $SD=1.02$). A post-hoc test found that day 2 ($M=3.78$, $SD=0.96$; *adj.* $p=0.025$) and day 3 were significantly higher ($M=3.92$, $SD=0.98$) in terms of trust compared to day 1 ($M=3.43$, $SD=0.99$; *adj.* $p<0.001$).

A post-hoc test found that day 3 was significantly higher ($M=4.05$, $SD=0.78$) in terms of predictability compared to day 1 ($M=3.83$, $SD=0.69$; *adj.* $p=0.040$).

4.2.2. Perceived safety - Table 1

A post-hoc test found that day 3 was significantly higher ($M=1.60$, $SD=1.66$) in terms of perceived safety compared to day 1 ($M=0.94$, $SD=1.70$; *adj.* $p=0.002$).

4.2.3. Driving style, longitudinal and lateral guidance - Table 2

Driving style was considered less safe in the morning ($M=2.35$, $SD=1.28$) than in the evening ($M=2.27$, $SD=1.22$).

Longitudinal guidance was perceived significantly worse in the morning ($M=5.59$, $SD=1.34$) than in the evening ($M=5.83$, $SD=1.32$).

4.2.4. Frequency of usage - Table 2

A post-hoc test found **no** significant differences on frequency of usage.

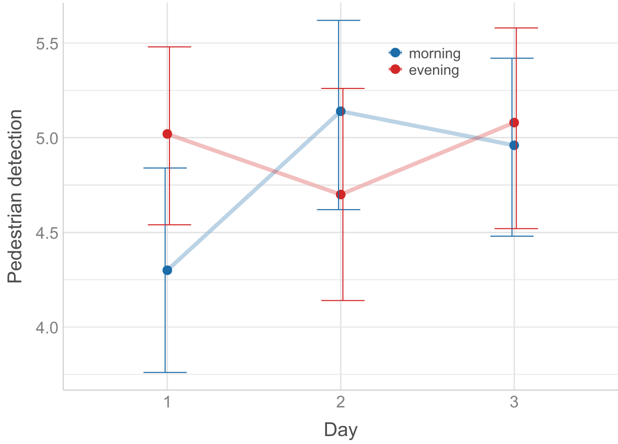

Fig. 7. Interaction effect of *Time of Day* × *Day* on pedestrian detection.

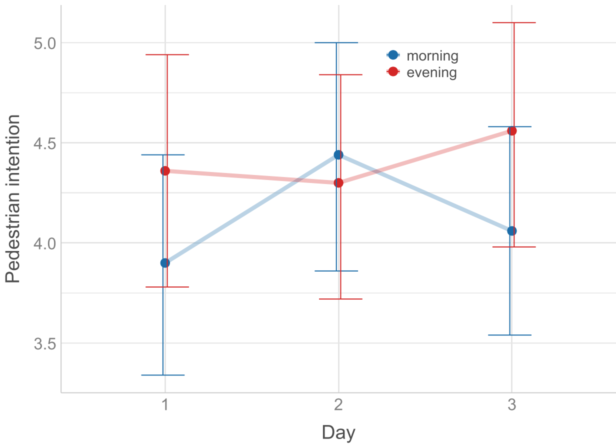

Fig. 8. Interaction effect of *Time of Day* × *Day* on pedestrian intention prediction.

Table 4
Statistics for all dependent variables (part 4/4).

Statistic	Unsafe Judgment	Appropriate Reaction to Env.	Own superiority of perf.	Clarity on what the AV will do next
Time of Day	$F(1,49) = 7.73$, $p = 0.008$	$F(1,49) = 5.91$, $p = 0.019$	$F(1,49) = 0.13$, $p = 0.72$	$F(1,49) = 13.62$, $p < 0.001$
Day	$F(2,98) = 0.87$, $p = 0.421$	$F(2,98) = 0.30$, $p = 0.74$	$F(2,98) = 0.27$, $p = 0.76$	$F(2,98) = 1.70$, $p = 0.19$
Interaction	$F(2,98) = 2.08$, $p = 0.13$	$F(2,98) = 1.56$, $p = 0.22$	$F(2,98) = 6.67$, $p = 0.002$	$F(2,98) = 3.13$, $p = 0.048$

4.2.5. AV capabilities: detection of pedestrians, vehicles, and signposts - Table 3

After initial skepticism, pedestrian detection later was higher and almost equal for morning and evening (see Fig. 7).

Participants perceived vehicle detection as significantly worse in the morning ($M = 5.68$, $SD = 1.44$) compared to the evening ($M = 5.84$, $SD = 1.42$).

A post-hoc test found no significant differences on signpost detection.

4.2.6. AV capabilities: prediction of pedestrians - Table 3

While the perceived capability of pedestrian prediction was almost equal for day 2 in the morning and evening, on days 1 and 3, the pedestrian intention prediction was perceived as significantly worse in the morning (see Fig. 8).

4.2.7. Unsafe judgment, appropriate reaction to environment, own superiority of performance, clarity on what the AV will do next - Table 4

The AV was perceived to have made significantly more unsafe judgments in the morning ($M = 2.67$, $SD = 1.89$) than in the evening ($M = 2.29$, $SD = 1.79$).

The appropriateness of reactions was lower in the morning ($M = 5.71$, $SD = 1.23$) than in the evening ($M = 5.91$, $SD = 1.23$).

Participants always believed to have performed better. This was higher in the evening than in the morning on day 1 but was inverted on day 3 (see Fig. 9).

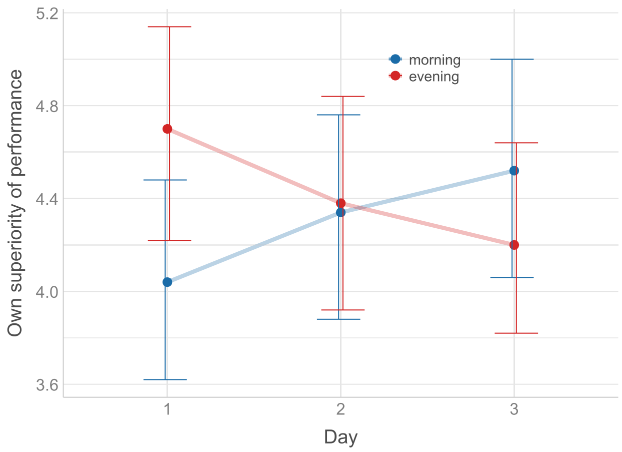

Fig. 9. Interaction effect of *Time of Day* × *Day* on own superiority of performance.

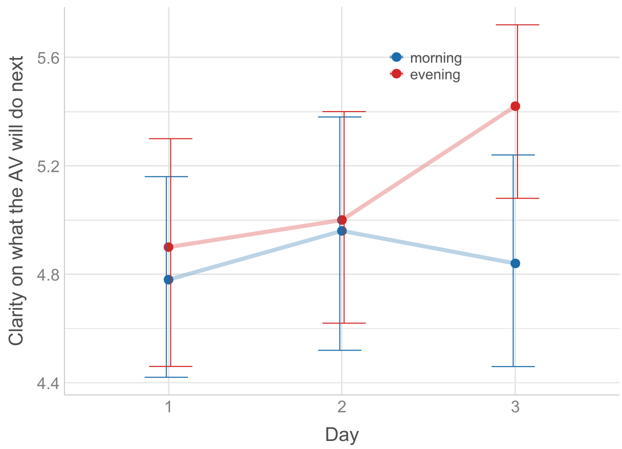

Fig. 10. Interaction effect of *Time of Day* × *Day* on clarity on what the AV will do next.

The clarity on what the AV will do nex was always rather high, with a small difference between morning and evening. However, on day 3, this difference became larger, with the evening being clearer (see Fig. 10).

4.2.8. Ranking

Finally, participants were asked to “Please sort the different visualizations according to your preferences.” The available items were: “Detection of Pedestrians (red coloring)”, “Detection of Vehicles (blue coloring)”, “Detection of Street Signs (yellow coloring)”, and “Intention of Pedestrians (blue icons above pedestrian heads)”.

A Friedman rank sum test found a significant effect of visualization on rank ($\chi^2(3) = 91.18$, $p = < 0.001$, r = 0.61; see Fig. 11). Vehicle detection was ranked as the most preferred.

4.3. Qualitative results

We conducted a thematic analysis (Braun & Clarke, 2006) on participants’ responses to the open-ended questions presented in Section 3.5. Providing feedback was voluntary. In total, 25 out of 50 participants responded. Two authors independently familiarized themselves with the data, generated inductive codes, and refined them into four overarching themes. The initial coding resulted in six discrepancies (approximately 8% of total codes). Intercoder reliability was almost perfect given Cohen’s $\kappa = 0.84$ (Landis & Koch, 1977). Discrepancies were discussed until a consensus was reached. Frequency counts indicate how many participants contributed to each theme. M1-M3 and E1-E3 denote participant responses from the first, second, and third morning and evening sessions, respectively, while “Final” indicates responses from the final three open-ended questions (see Section 3.5).

4.3.1. Theme 1: incorrect visualizations undermine trust (10/25)

Participants repeatedly described the AV’s pedestrian detection and intention overlays as unreliable, mirroring the documented inaccuracies of the underlying models (see Section 3.3). Rapidly flickering intention labels left several users anxious—“It had a lot

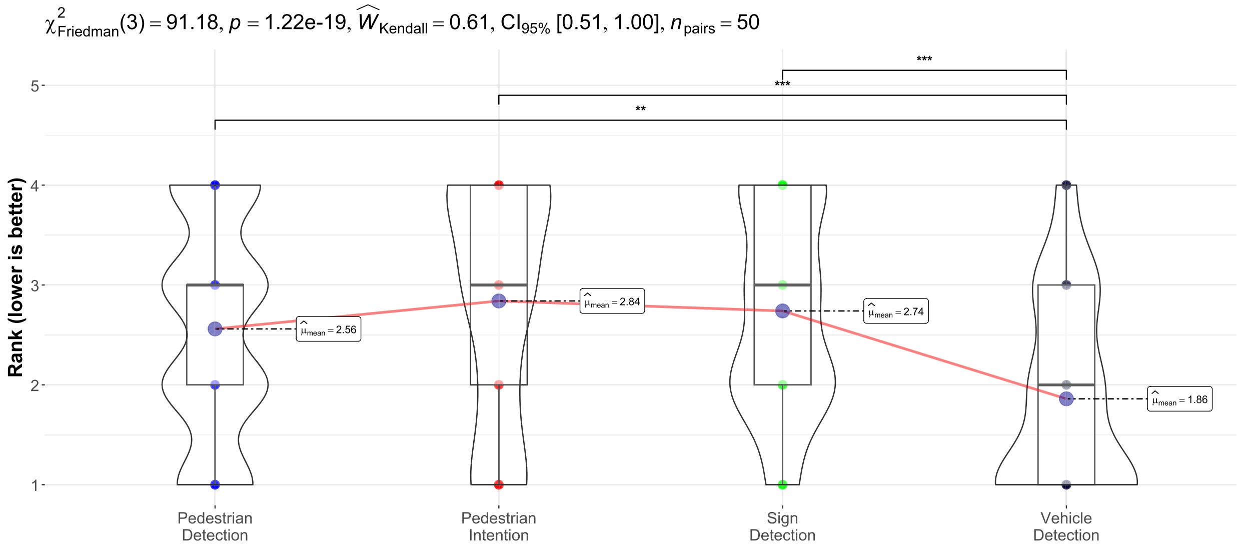

Fig. 11. Ranking of the different visualizations.

of trouble understanding the pedestrians' intents… it really made me anxious" [M1]; "It changed too much on a single pedestrian — I'm never sure if he is crossing or not" [E1]. Others pointed to outright misclassifications and misses: "A statue was indicated as a person" [Final]; "It missed a handful of people at a bus stop and didn't identify the guy on the moped until almost at the end" [E2]. Together, these comments show that participants read visualization quality as a direct signal of the AV's overall capability; when the overlays looked wrong, users relied on their own judgment, and their trust in the system eroded accordingly.

4.3.2. Theme 2: too many colors, too many flashes—unwanted effects of AV functionality visualizations (9/25)

Nine respondents felt that the color-coded overlays meant to clarify the driving scene instead produced visual clutter. Distinct hues for vehicles, signs, and pedestrians often conflicted with the real-world colors of traffic objects, turning the display into "color-coded chaos" that was "very distracting" [M1] and "confusing" when pedestrian icons changed rapidly [E1]. The problem was amplified by a strobe-like flickering: "The constant flickering of the icons above the pedestrians' heads was very distracting and incorrect more often than not," one participant noted, even reporting headaches [E2]. Several users called for simpler or adjustable visuals—"Maybe just highlight all the pedestrians so I know the car is registering them, and only mark those actually crossing" and "It would be good to turn up or turn down the level of detail displayed" [Final]. Together these comments show that, although the overlays faithfully expose AV processing, excessive colors and motion can undermine clarity and comfort, suggesting a need for streamlined or user-toggable displays.

4.3.3. Theme 3: smooth rides, but missing context—users want more AV feedback (4/25)

Even though the AV demonstrated smooth (i.e., without accidents or interrupting events) and realistic driving typical of daily commutes, many participants felt uneasy due to the lack of clear feedback on its intended actions. The absence of visible cues for turning, braking, and path planning led users to question the AV's overall reliability.

Several participants expressed needing more information about what the AV would do next. One participant commented, *"It would be interesting to have some kind of information regarding the path of the vehicle."* [Final] while another requested, *"I hope to know the signals of turning and braking to have a sense of safety."* [Final]. Furthermore, the desire for an immediate override was clear when one stated, *"I would want to know that I can immediately take over the wheel... if I had been in this car, I would have panicked."* [Final]. Despite the smooth ride, the lack of such real-time feedback undermines user trust.

4.3.4. Theme 4: gradual familiarity builds acceptance—but doubt grows over inconsistency (6/25)

Repeated morning-and-evening exposure helped some users move past their initial skepticism—"I do still find the indicators distracting, but I feel I am getting a bit more familiar" [M2]; "I am feeling safer as I gain trust in the AV" [E3]. Yet this emerging trust was fragile: new detection errors, especially in low light, quickly renewed concern. One participant observed, "At night the sign detection was very messy… a lot of lights got recognized as signs" [E2], while another noted, "The AV could not pick up on a large portion of lights, people, and bikes" [E3]. These recurring glitches reminded users of the system's limits, tempering growing acceptance and helping calibrate trust so it neither moves toward overtrust nor undertrust.

4.3.5. Theme 5: perception of AV benefits in traffic diminished by real-world complexity (8/25)

Participants recognized the substantial safety benefits promised by automated vehicles (AVs) yet simultaneously identified practical barriers that could negate these gains. Several respondents cited reduced accident risk, for example, "They will feel safer and less prone to accidents, such as those caused by drunk driving" [Final] and AVs "will save a lot of lives" [Final] but warned that public trust is fragile: “It takes just one story in the news for their reputation to be damaged” [Final], and mixed traffic with human-driven cars might “encourage reckless behavior.”

Comments also stressed that laboratory-style commutes are insufficient for judging real-world readiness. One participant asked, “How does the visual react during rainy days? Does it interfere with detecting people carrying umbrellas?” [Final], and requested trials involving emergency maneuvers or unexpected obstacles such as fallen trees. Another noted regional infrastructure differences: “I obtained my driver’s license in Canada, and the roads, traffic lights, and signs were slightly different than what I have experienced” [Final], implying that visualisation schemes must adapt to local norms. Collectively, these observations indicate that achieving the anticipated safety benefits will require robust AV performance across diverse environmental conditions and geographic contexts; otherwise, users’ optimism regarding AV technology could quickly turn to skepticism, and overreliance could become hazardous.

5. Discussion

This study explored the longitudinal effects of uncertainty visualization on user perceptions of AV functionality. Our findings indicate that sustained exposure to AV uncertainty visualizations may contribute to developing more realistic expectations of AV capabilities (despite making users also wary of them). These results expand upon previous work on trust in automation (Lee & See, 2004; Koo et al., 2015), suggesting that transparent communication of AV uncertainties can enhance trust without overwhelming users with mental workload (Kunze et al., 2018), especially also in the automotive domain (Colley et al., 2022, 2020, 2024). In the following, we discuss the implications of these findings, particularly regarding the potential of visualized uncertainties to foster calibrated trust and align user expectations with the realistic capabilities of AV technology. We also discuss practical applications for AV interface design and highlight future research to address study limitations and broaden our understanding of user trust in AVs.

5.1. The longitudinal effects of visualizing automation uncertainty

Prior research focused on short-term effects, frequently concluding that uncertainty visualizations enhance trust mainly when users experience repeated interactions with automation (Helldin et al., 2013). Our study extends these conclusions by answering the RQ “*What are the longitudinal effects of visualizing uncertainty of AVs’ Situation Detection and Situation Prediction in terms of users’ (1) trust, (2) perceived safety, (3) mental workload, (4) preference, (5) situation awareness, (6) perceived AV capability, and (7) acceptance?*”.

We found that longitudinal exposure to AV uncertainty visualizations significantly increases **trust** over multiple days (see Section 4.2.1). Specifically, participants reported significantly higher trust on day 3 than on day 1. This aligns with Frison et al. (2019), emphasizing the importance of consistent, verifiable information for trust formation in automation. Similarly, Lee and See (2004) highlighted iterative communication of AV decisions as critical for transparency, enhancing users’ trust over repeated interactions.

Qualitative data (see Section 4.3) further explained this relationship, highlighting some participants’ initial skepticism diminished gradually with increasing AV familiarity (Theme 4), reinforcing that repeated visualization interactions can improve trust. However, some participants noted barriers to trust improvements, such as ongoing pedestrian misclassification and unstable intention predictions (Theme 1). This aligns closely with Woide et al. (2022), arguing that trust only builds when visualized information consistently matches users’ real-world observations. Thus, ongoing visualization inaccuracies represent critical barriers to sustained trust improvements.

Additionally, **perceived safety** improved significantly by day 3 (see Section 4.2.2), indicating that uncertainty visualizations can improve participants’ perception of safety longitudinally. Some participants describe increasing feelings of safety alongside growing familiarity. This aligns with Ha et al. (2020) that explanations of AV behaviors increase perceived safety in low-risk situations. However, as noted in participants’ comments, certain inaccuracies (e.g., detecting pedestrians in the sky) triggered safety-related anxieties, aligning with Tennant et al. (2024), who identified concerns about unpredictable AV behavior as central to diminished perceived safety. These findings show that accurate and contextually appropriate visualizations are key to sustained perceived safety improvements.

Regarding **mental workload**, we found no significant changes over repeated exposures, consistent with prior work suggesting uncertainty visualizations need not increase mental workload if appropriately designed (Kunze et al., 2018). However, some participants’ comments reveal momentary discomfort and distraction from overly complex visualizations, such as rapid changes in pedestrian intention labels or excessive color use (Theme 2). This highlights the need for balancing visualization complexity with users’ mental capacities, in line with Colley et al. (2020), who found that clear and situated visualizations effectively reduce mental workload. We argue that, also in longitudinal contexts, situated visualizations should be preferred over textual explanations, see (Ha et al., 2020), as they reduce mental workload by eliminating the need for users to translate textual information into visual contexts mentally.

The visualization ranking (see Section 4.2.8) shows that participants **preferred** visualizations they perceived as consistently reliable (e.g., vehicle detection) over those seen as less reliable (e.g., pedestrian intention prediction). In the open-ended feedback, many of the 25 respondents criticized pedestrian intention icons due to frequent flickering and perceived unreliability, with some even suggesting their removal. Such criticisms suggest that some participants preferred not to see “under the hood,” particularly if internal uncertainties appeared confusing or seemingly non-functional. However, this preference does not justify omitting accurate visualizations of uncertainties; transparent representations remain necessary to calibrate user trust realistically. Nevertheless, previous work (Dixon, 2020; Jansen et al., 2024) strongly cautions against overly complex visualizations, as they may negatively impact user perceptions and impede proper trust calibration.

Besides, quantitative results of **situation awareness** reveal no significant changes longitudinally, indicating that sustained uncertainty visualizations neither significantly enhanced nor impaired situation awareness. However, qualitative results show some participants' demand for clearer information about upcoming AV actions, suggesting subjective gaps in perceived clarity despite the objective consistency of the AV's actions (Theme 3). This aligns with Omeiza et al. (2021), who reported that clear explanations improve users' functional understanding but do not necessarily lead to increased trust or improved situation awareness without additional context. Moreover, Schneider et al. (2021) similarly noted that while visualizations can enhance user experience, their influence on situation awareness is often limited.

Finally, **perceived AV capability** significantly increased for pedestrian and vehicle detection over time (see Sections 4.2.5 and 4.2.6). Yet, qualitative results reveal persistent concerns about visualization reliability in specific contexts (e.g., low lighting conditions or dynamic traffic), highlighting some participants' sensitivity to contextual factors in assessing AV capabilities (Themes 1 and 4). This aligns with Liu et al. (2022), who argued that misconceptions about AV capabilities could lead to inflated expectations, causing disappointment if misperformance. Similarly, Colley et al. (2021) argued that *genuine* visualizations (i.e., also showing misperformance) can improve users' understanding of AV capabilities but cautioned that improved understanding alone does not automatically translate into increased perceived AV capability.

Overall, **acceptance** toward AV technology improved through repeated visualization exposure. Participants reported gradually feeling safer and less anxious over multiple rides, consistent with prior suggestions that user acceptance evolves with prolonged experience and realistic expectations (Filiz, 2020; Ghazizadeh et al., 2012). Nevertheless, some participants noted intermittent inaccuracies, such as erroneous pedestrian detections at night, which reminded them of ongoing AV limitations (Theme 4). This finding underscores the importance of context-accurate visualizations for sustained acceptance.

Although the time-of-day effects were statistically significant, the absolute gaps were small. For example, trust increased by only 0.04 points (3.69 vs. 3.73) in the evening, constituting a neutral trust rating. Moreover, other significant measures increased by less than 0.25 points in the evening (e.g., longitudinal guidance 5.59 vs. 5.83; unsafe judgments 2.67 vs. 2.29). These minor yet consistent improvements show that these measures are subtly context-dependent. Context is, for example, the time of day and environmental lighting (see Theme 5). Thus, we argue that AV uncertainty visualization designs should be context-adaptive, aligning with the suggestions of Jansen et al. (2025). Future research should investigate how context, such as dynamic traffic environments (urban, rural, or highway), influence users' longitudinal perceptions of AVs.

5.2. Influencing the public perception of automated vehicles

Our findings underline the importance of informing the public about AV uncertainties, aligning with calls for transparent AV interfaces (Ekman et al., 2018). Previous research identified perceived AV capability, trust, and perceived safety as central factors influencing public acceptance and adoption of AVs (Ghazizadeh et al., 2012; Liu et al., 2022; Tennant et al., 2024). Our results support these arguments by demonstrating that uncertainty visualizations effectively align participants' expectations with realistic AV capabilities, reducing misconceptions and inflated expectations regarding AV readiness and thus mitigating safety risks associated with overreliance on automation (Filiz, 2020). Participants who answered our open-ended questions emphasized a clear preference for simplified and accurate visualizations over complex iconography, which often led to confusion. This preference underscores AV manufacturers' need to avoid "autonowashing" (Dixon, 2020) and adopt interfaces that realistically communicate system limitations (Jansen et al., 2024). To further enhance acceptance, uncertainty visualizations could be personalized to individual preferences and situational needs, for instance, through human-in-the-loop optimization methods as demonstrated by Jansen et al. (2025).

Furthermore, our qualitative results reveal that some participants connected uncertainty visualizations to broader public discussions about AV safety and acceptance. While many participants acknowledged the potential societal benefits of AVs, such as reduced accidents from human error, they also expressed concerns regarding overreliance on automated technology (Theme 5). Participants pointed out the risk of complacency among drivers, indicating that if uncertainty is not communicated transparently, they might prematurely assume AV readiness, which aligns with warnings by Liu et al. (2022) and Tennant et al. (2024). Transparent uncertainty visualizations could, therefore, act as safeguards against public misconceptions and help balance enthusiasm for AV benefits with realistic awareness of current limitations.

Also, some participants questioned AV performance in scenarios beyond routine commutes and stressed that public trust in AV technology requires evidence of reliable performance across diverse and challenging situations (Theme 5). This aligns with Jing et al. (2023) and Du et al. (2019), who argued that public communication about AV technology often neglects challenging operational scenarios, potentially leading to overly optimistic expectations. Thus, transparently visualizing uncertainty in challenging conditions can substantially influence public perception by realistically calibrating user expectations of AVs.

5.3. Practical implications

The practical implications of our study are multifaceted. First, the increased trust over repeated interactions with uncertainty visualizations suggests that longitudinal exposure to AV functionalities could be an effective strategy for trust-building in commercial AV applications. This finding is particularly relevant for scenarios where repeated interactions are common, such as ride-hailing services or daily commuting contexts. Consistent exposure to AV technology over extended periods may help users gradually calibrate their trust, potentially explaining anecdotal reports of Waymo users maintaining loyalty to the service.

Second, our findings strongly argue for simplicity and clarity in visualizations, reflecting participants' consistent preferences for clear, color-based uncertainty cues over complex or rapidly changing visual representations (Theme 2). This aligns with recommendations by Kunze et al. (2018). It is further supported by Colley et al. (2020) and Currano et al. (2021), who found intuitive and minimalistic visualizations most effective in enhancing user experience and reducing cognitive load. Therefore, AV interface designers should prioritize visualizations that communicate essential uncertainties without unnecessary complexity.

Third, some participants emphasized the importance of feedback regarding the AV's intended actions (Theme 3). Although our quantitative results suggested generally consistent clarity about AV actions (see Section 4.2.7), some participants textual feedback argued for clearer, real-time communication of imminent actions, such as braking or turning signals. These findings resonate with previous work emphasizing that explicit explanatory information helps build trust and perceived safety, particularly in dynamic driving scenarios (Koo et al., 2015; Ha et al., 2020). Therefore, in addition to Situation Detection and Situation Prediction, future work should also visualize predictive feedback on upcoming AV maneuvers—commonly known as Trajectory Planning in the AV functional hierarchy (Dietmayer, 2016). This may include head-up displays that show the vehicle's dashboard information and planned driving paths (Jansen et al., 2025; Colley et al., 2021). Investigating these visualizations longitudinally would provide insights into how predictive trajectory feedback affects user perceptions of AV technology over time.

5.4. Limitations and future work

While our longitudinal study design provides insights into the effects of sustained exposure to AV uncertainty visualizations, several limitations exist. Our study's reliance on video-based simulations rather than fully immersive environments like real AVs may limit the ecological validity of our findings. This concern aligns with discussions in the literature about the importance of ecological validity in simulation studies (Pradhan et al., 2019). Future research should replicate our study in real or fully interactive simulated AV environments, allowing participants to experience physical cues (e.g., motion) and vehicle dynamics (Sasalovici et al., 2025). These physical cues could significantly enhance findings' ecological validity regarding trust dynamics (Stange et al., 2022).

Additionally, the study sample was predominantly composed of younger participants due to the nature of our online recruitment platform, Prolific. This sample bias is a common issue in online research and can affect the generalizability of the results. Future studies should deliberately target a broader and more representative demographic, including older adults, who exhibit distinct patterns in trust and interaction with AV technologies (Haghzare et al., 2021). A larger and more diverse sample would improve our understanding of demographic-specific responses to AV uncertainty visualization.

Our study assigned sessions to "morning" and "evening" blocks yet let participants begin at any convenient time within those windows. This led to a 12 pm session being closer to a 4 pm session than to a 7 am session. While the scheduling flexibility eased participant recruitment, it introduced intra-block variation that may have diluted time-of-day effects. Future research could narrow the windows or enforce fixed intervals between rides to examine temporal factors more precisely.

Moreover, survey fatigue may have affected our study's participant engagement and data quality. Although each session lasted only about 14 minutes (on average 81 minutes across three days) and we included attention checks to monitor engagement, future research should incorporate quantitative measures of fatigue, such as self-report scales or physiological indicators, which could help quantify any fatigue-related impacts.

Another limitation is the study duration, as we examined effects across only three consecutive days. A recent systematic review of 96 driver-vehicle studies found that research already labels exposures of two sessions or more as "long-term," concluding that changes in users' perceptions depend more on repeated interactions than on absolute calendar length (Nkusi et al., 2025). Consistent with that insight, our three-day (six-session) procedure can already reveal significant changes in trust, perceived safety, and capability ratings, mirroring other longitudinal studies (Colley et al., 2024, 2022; Ulahannan et al., 2021). Yet, slower changes may unfold over weeks or months. Future work should, therefore, extend exposure periods.

Lastly, voluntary, open-ended questions may have limited both the number and depth of qualitative responses. Overall, 25 of the 50 participants provided written comments, creating a self-selected subset whose willingness to invest extra effort likely reflects greater interest and motivation in AV topics. Future work could offer incentives, embedding brief reflection breaks within the task flow, or conducting structured one-on-one interviews to elicit equally thorough feedback from a broader cross-section of participants. Additionally, follow-up studies might compare responses from motivated volunteers to those from randomly sampled users to gauge the robustness of the qualitative themes.

6. Conclusion

Our longitudinal study with $N = 50$ participants suggests that sustained exposure to AV uncertainty visualizations can help align user expectations more closely with actual AV capabilities. Through transparency, AV systems may better balance simplicity and informative feedback to enhance perceived safety and public acceptance. These findings underscore that clear communication of AV limitations is essential for encouraging realistic user expectations, which is critical for safe AV adoption.

Looking ahead, future research should extend to real-world AV interactions to fully understand how users' trust evolves with repeated, "hands-on" exposure. Such insights will guide AV designers in developing responsible, user-centered systems that support broader societal readiness for automated driving.

CRediT authorship contribution statement

Pascal Jansen: Writing – review & editing, Writing – original draft, Visualization, Software, Project administration, Methodology, Investigation, Data curation, Conceptualization. **Mark Colley:** Writing – review & editing, Writing – original draft, Visualization, Methodology, Formal analysis, Data curation, Conceptualization. **Max Rädler:** Writing – review & editing, Writing – original draft, Visualization, Software, Methodology, Investigation, Data curation, Conceptualization. **Jonas Schwedler:** Software, Methodology, Investigation. **Enrico Rukzio:** Writing – review & editing, Supervision, Resources.

Open science

We open-sourced[5] the code of our longitudinal study website, which was used to host the videos and track participants' progress. We will make the study data available upon acceptance.

Declaration of generative AI and AI-assisted technologies in the writing process

During the preparation of this manuscript, the authors utilized Grammarly and ChatGPT to enhance grammar and correct spelling. After employing these tools, the authors thoroughly reviewed and edited the content as necessary, assuming full responsibility for the integrity and accuracy of the published work.

Acknowledgements

We thank all study participants.

Appendix A. Supplementary material

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.trf.2025.05.013.

Data availability

Data will be made available on request.

References

Arnab, A., & Torr, P. H. S. (2017). Pixelwise instance segmentation with a dynamically instantiated network. arXiv:1704.02386.

Beller, J., Heesen, M., & Vollrath, M. (2013). Improving the driver–automation interaction: An approach using automation uncertainty. *Human Factors*, *55*, 1130–1141. PMID: 24745204.

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, *3*, 77–101.

Caine, K. (2016). Local standards for sample size at chi. In *Proceedings of the 2016 CHI conference on human factors in computing systems, CHI '16* (pp. 981–992). New York, NY, USA: Association for Computing Machinery.

Colley, M., Bräuner, C., Lanzer, M., Walch, M., Baumann, M., & Rukzio, E. (2020). Effect of visualization of pedestrian intention recognition on trust and cognitive load. In *12th international conference on automotive user interfaces and interactive vehicular applications, AutomotiveUI '20* (pp. 181–191). New York, NY, USA: Association for Computing Machinery.

Colley, M., Eder, B., Rixen, J. O., & Rukzio, E. (2021). Effects of semantic segmentation visualization on trust, situation awareness, and cognitive load in highly automated vehicles. In *Proceedings of the 2021 CHI conference on human factors in computing systems*. New York, NY, USA: Association for Computing Machinery.

Colley, M., Krauss, S., Lanzer, M., & Rukzio, E. (2021). How should automated vehicles communicate critical situations? A comparative analysis of visualization concepts. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, *5*.

Colley, M., Bajrovic, E., & Rukzio, E. (2022). Effects of pedestrian behavior, time pressure, and repeated exposure on crossing decisions in front of automated vehicles equipped with external communication. In *Proceedings of the 2022 CHI conference on human factors in computing systems* (pp. 1–11).

Colley, M., Rädler, M., Glimmann, J., & Rukzio, E. (2022). Effects of scene detection, scene prediction, and maneuver planning visualizations on trust, situation awareness, and cognitive load in highly automated vehicles. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, *6*.

Colley, M., Kornmüller, D., Dey, D., Ju, W., & Rukzio, E. (2024). Longitudinal effects of external communication of automated vehicles in the usa and Germany: A comparative study in virtual reality and via a browser. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, *8*.

Colley, M., Speidel, O., Strohbeck, J., Rixen, J. O., Belz, J. H., & Rukzio, E. (2024). Effects of uncertain trajectory prediction visualization in highly automated vehicles on trust, situation awareness, and cognitive load. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, *7*.

Cordts, M., Omran, M., Ramos, S., Rehfeld, T., Enzweiler, M., Benenson, R., Franke, U., Roth, S., & Schiele, B. (2016). The cityscapes dataset for semantic urban scene understanding. In *2016 IEEE conference on computer vision and pattern recognition (CVPR)* (pp. 3213–3223). Las Vegas, NV, USA: IEEE.

Currano, R., Park, S. Y., Moore, D. J., Lyons, K., & Sirkin, D. (2021). Little road driving hud: Heads-up display complexity influences drivers' perceptions of automated vehicles. In *Proceedings of the 2021 CHI conference on human factors in computing systems*. New York, NY, USA: Association for Computing Machinery.

Dietmayer, K. (2016). Predicting of machine perception for automated driving. In *Autonomous driving* (pp. 407–424). Berlin Heidelberg, Berlin, Heidelberg: Springer.

Dixon, L. (2020). Autonowashing: The greenwashing of vehicle automation. *Transportation Research Interdisciplinary Perspectives*, *5*, Article 100113.

Du, N., Haspiel, J., Zhang, Q., Tilbury, D., Pradhan, A. K., Yang, X. J., & Robert, L. P. (2019). Look who's talking now: Implications of av's explanations on driver's trust, av preference, anxiety and mental workload. *Transportation Research. Part C, Emerging Technologies*, *104*, 428–442. ID: 271729.

Ekman, F., Johansson, M., & Sochor, J. (2018). Creating appropriate trust in automated vehicle systems: A framework for hmi design. *IEEE Transactions on Human-Machine Systems*, *48*, 95–101.

Faas, S. M., Kao, A. C., & Baumann, M. (2020). A longitudinal video study on communicating status and intent for self-driving vehicle – pedestrian interaction. In *Proceedings of the 2020 CHI conference on human factors in computing systems, CHI '20* (pp. 1–14). New York, NY, USA: Association for Computing Machinery.

Fagnant, D. J., & Kockelman, K. (2015). Preparing a nation for autonomous vehicles: Opportunities, barriers and policy recommendations. *Transportation Research. Part A, Policy and Practice*, *77*, 167–181.

Faul, F., Erdfelder, E., Buchner, A., & Lang, A.-G. (2009). Statistical power analyses using g* power 3.1: Tests for correlation and regression analyses. *Behavior Research Methods*, *41*, 1149–1160.

[5] https://github.com/Pascal-Jansen/longitudinal-uncertainty-visualization.

Filiz, C. (2020). Can autonomous vehicles prevent traffic accidents? In *Accident analysis and prevention*. Rijeka, Croati: IntechOpen.
Frison, A.-K., Wintersberger, P., Riener, A., Schartmüller, C., Boyle, L. N., Miller, E., & Weigl, K. (2019). In ux we trust: Investigation of aesthetics and usability of driver-vehicle interfaces and their impact on the perception of automated driving. In *Proceedings of the 2019 CHI conference on human factors in computing systems, CHI '19* (pp. 1–13). New York, NY, USA: Association for Computing Machinery.
Ghazizadeh, M., Lee, J. D., & Boyle, L. N. (2012). Extending the technology acceptance model to assess automation. *Cognition Technology & Work*, *14*, 39–49.
Girshick, R., Donahue, J., Darrell, T., & Malik, J. (2014). Rich feature hierarchies for accurate object detection and semantic segmentation. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition (CVPR)*.
Ha, T., Kim, S., Seo, D., & Lee, S. (2020). Effects of explanation types and perceived risk on trust in autonomous vehicles. *Transportation Research. Part F, Traffic Psychology and Behaviour*, *73*, 271–280.
Haghzare, S., Campos, J. L., Bak, K., & Mihailidis, A. (2021). Older adults' acceptance of fully automated vehicles: Effects of exposure, driving style, age, and driving conditions. *Accident Analysis and Prevention*, *150*, Article 105919.
Hart, S. G., & Staveland, L. E. (1988). *Development of nasa-tlx (task load index): Results of empirical and theoretical research*. *Advances in psychology: Vol. 52*. Amsterdam, the Netherlands: Elsevier (pp. 139–183).
Helldin, T., Falkman, G., Riveiro, M., & Davidsson, S. (2013). Presenting system uncertainty in automotive uis for supporting trust calibration in autonomous driving. In *Proceedings of the 5th international conference on automotive user interfaces and interactive vehicular applications, AutomotiveUI '13* (pp. 210–217). New York, NY, USA: Association for Computing Machinery.
Hilgarter, K., & Granig, P. (2020). Public perception of autonomous vehicles: A qualitative study based on interviews after riding an autonomous shuttle. *Transportation Research. Part F, Traffic Psychology and Behaviour*, *72*, 226–243.
International, S. A. E. (2023). Sae levels of driving automation™ refined for clarity and international audience. https://www.sae.org/blog/sae-j3016-update.
Jansen, P., Colley, M., & Rukzio, E. (2022). A design space for human sensor and actuator focused in-vehicle interaction based on a systematic literature review. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, *6*.
Jansen, P., Colley, M., Pfeifer, T., & Rukzio, E. (2024). Visualizing imperfect situation detection and prediction in automated vehicles: Understanding users' perceptions via user-chosen scenarios. *Transportation Research. Part F, Traffic Psychology and Behaviour*, *104*, 88–108.
Jansen, P., Colley, M., Krauß, S., Hirschle, D., & Rukzio, E. (2025). Opticarvis: Improving automated vehicle functionality visualizations using Bayesian optimization to enhance user experience. In *Proceedings of the 2025 CHI conference on human factors in computing systems*. New York, NY, USA: Association for Computing Machinery.
Jing, P., Cai, Y., Wang, B., Wang, B., Huang, J., Jiang, C., & Yang, C. (2023). Listen to social media users: Mining Chinese public perception of automated vehicles after crashes. *Transportation Research. Part F, Traffic Psychology and Behaviour*, *93*, 248–265.
Koo, J., Kwac, J., Ju, W., Steinert, M., Leifer, L., & Nass, C. (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*, *9*, 269–275.
Körber, M. (2019). Theoretical considerations and development of a questionnaire to measure trust in automation. In S. Bagnara, R. Tartaglia, S. Albolino, T. Alexander, & Y. Fujita (Eds.), *Proceedings of the 20th congress of the international ergonomics association (IEA 2018)* (pp. 13–30). Cham: Springer International Publishing.
Kothgassner, O., Felnhofer, A., Hauk, N., Kastenhofer, E., Gomm, J., & Krysprin-Exner, I. (2013). *Technology usage inventory. Manual*. *Icarus*, *17*, 90. (Online; accessed: 05-July-2020).
Kreiss, S., Bertoni, L., & Alahi, A. (2021). OpenPifPaf: Composite fields for semantic keypoint detection and spatio-temporal association. *IEEE Transactions on Intelligent Transportation Systems*, 1–14.
Kunze, A., Summerskill, S. J., Marshall, R., & Filtness, A. J. (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, AutomotiveUI '18* (pp. 164–175). New York, NY, USA: Association for Computing Machinery.
Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, 159–174.
Large, D. R., Burnett, G., Morris, A., Muthumani, A., & Matthias, R. (July 17). A longitudinal simulator study to explore drivers' behaviour during highly-automated driving. In *Advances in human aspects of transportation: Proceedings of the AHFE 2017 international conference on human factors in transportation* (pp. 583–594). Springer.
Lee, J. D., & See, K. A. (2004). Trust in automation: Designing for appropriate reliance. *Human Factors*, *46*, 50–80.
Leon, F., & Gavrilescu, M. (2021). A review of tracking and trajectory prediction methods for autonomous driving. *Mathematics*, *9*.
Lindemann, P., Lee, T.-Y., & Rigoll, G. (2018). Catch my drift: Elevating situation awareness for highly automated driving with an explanatory windshield display user interface. *Multimodal Technologies and Interaction*, *2*, 71.
Liu, P., Du, M., Xu, Z., & Chu, Y. (2022). People with more misconceptions about automated vehicles might be more positive toward them. *Transportation Research. Part F, Traffic Psychology and Behaviour*, *87*, 264–278.
Mordan, T., Cord, M., Pérez, P., & Alahi, A. (2021). Detecting 32 pedestrian attributes for autonomous vehicles. *IEEE Transactions on Intelligent Transportation Systems*, *23*, 1–13.
Muir, B. M., & Moray, N. (1996). Trust in automation. Part ii. Experimental studies of trust and human intervention in a process control simulation. *Ergonomics*, *39*, 429–460.
Nkusi, E. M., Grabbe, N., & Bengler, K. (2025). Long-term is no term: A systematic review of learning effects and the varied use of "long-term" in driver–vehicle interaction. *Transportation Research. Part F, Traffic Psychology and Behaviour*, *112*, 111–137.
Omeiza, D., Kollnig, K., Web, H., Jirotka, M., & Kunze, L. (2021). Why not explain? Effects of explanations on human perceptions of autonomous driving. In *2021 IEEE international conference on advanced robotics and its social impacts (ARSO)* (pp. 194–199). New York, NY, USA: IEEE.
Peintner, J. B., Manger, C., & Riener, A. (2022). "can you rely on me?" evaluating a confidence hmi for cooperative, automated driving. In *Proceedings of the 14th international conference on automotive user interfaces and interactive vehicular applications, AutomotiveUI '22* (pp. 340–348). New York, NY, USA: Association for Computing Machinery.
Pfleging, B., Rang, M., & Broy, N. (2016). Investigating user needs for non-driving-related activities during automated driving. In *Proceedings of the 15th international conference on mobile and ubiquitous multimedia, MUM '16* (pp. 91–99). New York, NY, USA: Association for Computing Machinery.
Pradhan, A., Jeong, H., & Ross, B. (2019). Is driving simulation a viable method for examining drivers' ethical choices? An exploratory study. In *Proceedings of the driving assessment conference: Vol. 10*.
Rasouli, A., & Tsotsos, J. K. (2019). Autonomous vehicles that interact with pedestrians: A survey of theory and practice. *IEEE Transactions on Intelligent Transportation Systems*, *21*, 900–918.
Rasouli, A., Kotseruba, I., & Tsotsos, J. K. (2017). Are they going to cross? A benchmark dataset and baseline for pedestrian crosswalk behavior. In *Proceedings of the IEEE international conference on computer vision workshops* (pp. 206–213). IEEE.
Sasalovici, M., Zeqiri, A., Schramm, R. C., Nunez, O. J. A., Jansen, P., Freiwald, J. P., Colley, M., Winkler, C., & Rukzio, E. (2025). Bumpy Ride? Understanding the Effects of External Forces on Spatial Interactions in Moving Vehicles. In *Proceedings of the 2025 CHI conference on human factors in computing systems*. New York, NY, USA: Association for Computing Machinery.
Schneider, T., Hois, J., Rosenstein, A., Ghellal, S., Theofanou-Fülbier, D., & Gerlicher, A. R. (2021). Explain yourself! Transparency for positive ux in autonomous driving. In *Proceedings of the 2021 CHI conference on human factors in computing systems*. New York, NY, USA: Association for Computing Machinery.

Schneider, T., Hois, J., Rosenstein, A., Metzl, S., Gerlicher, A. R., Ghellal, S., & Love, S. (2023). Don't fail me! The level 5 autonomous driving information dilemma regarding transparency and user experience. In *Proceedings of the 28th international conference on intelligent user interfaces, IUI '23* (pp. 540–552). New York, NY, USA: Association for Computing Machinery.

Stange, V., Goralzik, A., Ernst, S., Steimle, M., Maurer, M., & Vollrath, M. (2022). Please stop now, automated vehicle!–passengers aim to avoid risk experiences in interactions with a crossing vulnerable road user at an urban junction. *Transportation Research. Part F, Traffic Psychology and Behaviour*, *87*, 164–188.

Taylor, R. M. (2017). Situational awareness rating technique (sart): The development of a tool for aircrew systems design. In *Situational awareness* (pp. 111–128). Abingdon, UK: Routledge.

Tennant, C., Stilgoe, J., Vucevic, S., & Stares, S. (2024). Public anticipations of self-driving vehicles in the uk and us. *Mobilities*, 1–18.

Ulahannan, A., Thompson, S., Jennings, P., & Birrell, S. (2021). Using glance behaviour to inform the design of adaptive HMI for partially automated vehicles. *IEEE Transactions on Intelligent Transportation Systems*, *23*, 4877–4892.

Van Der Laan, J. D., Heino, A., & De Waard, D. (1997). A simple procedure for the assessment of acceptance of advanced transport telematics. *Transportation Research. Part C, Emerging Technologies*, *5*, 1–10.

Venkatesh, V., Morris, M. G., Davis, G. B., & Davis, F. D. (2003). User acceptance of information technology: Toward a unified view. *MIS Quarterly*, 425–478.

Wang, W., Dai, J., Chen, Z., Huang, Z., Li, Z., Zhu, X., Hu, X., Lu, T., Lu, L., Li, H., et al. (2023). Internimage: Exploring large-scale vision foundation models with deformable convolutions. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (pp. 14408–14419). New York, NY, USA: IEEE.

Wilbrink, M., Schieben, A., & Oehl, M. (2020). Reflecting the automated vehicle's perception and intention: Light-based interaction approaches for on-board hmi in highly automated vehicles. In *Proceedings of the 25th international conference on intelligent user interfaces companion, IUI '20* (pp. 105–107). New York, NY, USA: Association for Computing Machinery.

Wintersberger, P., Frison, A.-K., Riener, A., & Sawitzky, T. v. (2019). Fostering user acceptance and trust in fully automated vehicles: Evaluating the potential of augmented reality. *PRESENCE: Virtual and Augmented Reality*, *27*, 46–62.

Wobbrock, J. O., Findlater, L., Gergle, D., & Higgins, J. J. (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, CHI '11* (pp. 143–146). New York, NY, USA: Association for Computing Machinery.

Woide, M., Colley, M., Damm, N., & Baumann, M. (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, AutomotiveUI '22* (pp. 119–130). New York, NY, USA: Association for Computing Machinery.