Highlights

Visualizing Imperfect Situation Detection and Prediction in Automated Vehicles: Understanding Users' Perceptions via User-Chosen Scenarios

Pascal Jansen,Mark Colley,Tim Pfeifer,Enrico Rukzio

- Introduction of EDULICIT, a novel method for eliciting user-centric driving scenarios while educating the public on the functionalities and challenges of automated vehicles.
- Application of state-of-art neural networks on videos of driving scenarios to visualize current situation detection, prediction, and trajectory planning functionalities of automated vehicles.
- Comparison to previous work showing that our approach replicates it.
- Demonstration of EDULICIT through a website implementation enabling users to upload driving scenario videos they deem challenging for automated vehicles in the wild. The website automatically annotates these videos with the functionalities visualizations to educate the general public about automated vehicle functionalities.
- User-chosen driving scenarios primarily include large intersections and/or multiple vulnerable road users.
- Users perceive vulnerable road users as more unreliable, unpredictable, and harder to detect by sensors than vehicles.

Visualizing Imperfect Situation Detection and Prediction in Automated Vehicles: Understanding Users' Perceptions via User-Chosen Scenarios

Pascal Jansen[a,*], Mark Colley[a,*], Tim Pfeifer[a] and Enrico Rukzio[a]

[a]*Institute of Media Informatics Ulm University Ulm Germany*

ARTICLE INFO

Keywords:
automated vehicles
automotive
explainable
artificial intelligence

ABSTRACT

User acceptance is essential for successfully introducing automated vehicles (AVs). Understanding the technology is necessary to overcome skepticism and achieve acceptance. This could be achieved by visualizing (uncertainties of) AV's internal processes, including situation perception, prediction, and trajectory planning. At the same time, relevant scenarios for communicating the functionalities are unclear. Therefore, we developed EDULICIT to concurrently elicit relevant scenarios and evaluate the effects of visualizing AV's internal processes. A website capable of showing annotated videos enabled this methodology. With it, we replicated the results of a previous online study (N=76) using pre-recorded real-world videos. Additionally, in a second online study (N=22), participants uploaded scenarios they deemed challenging for AVs using our website. Most scenarios included large intersections and/or multiple vulnerable road users. Our work helps assess scenarios perceived as challenging for AVs by the public and, simultaneously, can help educate the public about visualizations of the functionalities of current AVs.

1. Introduction

With automated vehicles (AVs), users can engage in non-driving activities such as reading, working, or even sleeping (Pfleging, Rang and Broy, 2016). However, with AVs at Society of Automotive Engineers (SAE) levels 4 or 5 controlling the driving task and thus managing safety-critical maneuvers, challenges arise concerning both overtrust and undertrust among users. For example, Tennant, Stilgoe, Vucevic and Stares (2024) found that 65% of respondents expressed safety concerns related to unreliable AV behavior, such as malfunctions. Hilgarter and Granig (2020) also report that potential users worry about the AVs' reliability, which refers to the AV's ability to act correctly. With high reliability, if users do not trust the AV, undertrust is prevalent. Uncertainty refers to the inverse of reliability. With high reliability, there is low uncertainty in successful task completion and vice versa. This shows that many potential users might not use AVs due to potential undertrust. While AVs could reduce accidents (Filiz, 2020; Fagnant and Kockelman, 2015) and enable novel activities such as reading or sleeping, they will only be used with sufficient trust. Liu, Du, Xu and Chu (2022) also showed that there are numerous misconceptions in the public and that people with these misconceptions (i.e., "AVs are already available in the market," "AVs do not need to be driven manually at all," and "Mature business models for AVs have been established.") are more likely to have positive attitudes toward AVs (i.e., show overtrust). This shows that **user perceptions** of AVs are flawed. In this context, additional indicators of user perceptions are trust, perceived safety, and the user's assessment of the AV's ability to perceive the surroundings and conduct the driving task, which we focused on in this work.

One possibility to appropriate user perceptions could be communicating AV functionality to the potential user. This functionality can be broadly defined into Situation Detection (i.e., detecting relevant objects), Situation Prediction (i.e., predicting future positions and states of relevant objects), and Trajectory planning (i.e., planning the AV's future movements). Therefore, for example, prior work visualized the detection of other vehicles (by highlighting) in poor weather circumstances (Wintersberger, Frison, Riener and Sawitzky, 2019), visualizing the derived pedestrian intention (Colley, Bräuner, Lanzer, Walch, Baumann and Rukzio, 2020), and general object detection (e.g., pedestrians,

*Joint First Authors

pascal.jansen@uni-ulm.de (P. Jansen); mark.colley@uni-ulm.de (M. Colley); tim.pfeifer@uni-ulm.de (T. Pfeifer); enrico.rukzio@uni-ulm.de (E. Rukzio)
ORCID(s): 0000-0002-9335-5462 (P. Jansen); 0000-0001-5207-5029 (M. Colley); 0000-0002-6387-3621 (T. Pfeifer); 0000-0002-4213-2226 (E. Rukzio)

signposts, bicyclists, and vehicles) by showing the results of the semantic segmentation (Colley, Eder, Rixen and Rukzio, 2021a). Colley, Rädler, Glimmann and Rukzio (2022) combined some of these visualizations and included vehicle trajectories for the ego and other vehicles. On the other hand, overtrust encourages the user's overuse and undermonitoring of the AV and the potential to misuse such systems.

While these visualizations affected trust and also conveyed AV reliability to the user, some works were done with a limited number of participants (Colley et al., 2020; Wintersberger et al., 2019) or with a very specific use case (Wintersberger et al., 2019; Colley et al., 2022, 2021a). Therefore, the generalizability of these results is limited. Additionally, with the recent practice "autonowashing" (Dixon, 2020), that is, overstating the functionalities of vehicles, there is a need to inform future users and the public, that is, the entirety of potential future users of AVs about AVs' actual functionalities.

As the literature showed, providing visualizations can be leveraged to present the functionalities of AVs. Providing this information in an easily accessible manner, for example, via an online website, could alleviate the need to educate the public about AVs' functionalities to reduce potential side effects of "autonowashing" (Dixon, 2020) to calibrate expectations towards AVs. However, such an approach is currently missing, and vehicle manufacturers are unlikely to communicate their AV software's weaknesses. Thus, an alternative is to show state-of-the-art academic research in the relevant areas (e.g., Situation Detection).

In this work, we close this gap by providing interested individuals and, through this, the general public with a website that visualizes the current functionalities of open-source approaches regarding Situation Detection and Situation Prediction. This website constitutes a novel method called EDULICIT to elicit scenarios that potential users of AVs might consider dangerous or challenging to AVs. These scenarios are elicited by allowing users to upload scenarios they find challenging for potential AVs. This provides insights for researchers evaluating solutions to overcome, for example, the initial undertrust regarding AVs. We implemented a website capable of (1) showing the videos employed by Colley et al. (2022) that show an AV driving through a lively city in Germany and (2) automatically annotating uploaded videos with Situation Detection and Situation Prediction information (see Figure 1). We conducted a user study with this website and evaluated the results of visualizing functionalities and the scenarios participants deemed challenging. This work presents a mixed-methods approach. We employ a public website with automatic video processing as a method to democratize access to AV feedback visualizations. Additionally, we use qualitative methods to evaluate the scenarios uploaded by participants (which they deemed challenging for AVs) and quantitatively analyze the subjective impact of visual annotations in the videos.

Contribution Statement: (1) The method EDULICIT to simultaneously present the public current functionalities of detecting objects (Situation Detection) and predicting the pedestrian crossing decisions (staying on the sidewalk, crossing, unsure; Situation Prediction) in the immediate future while eliciting scenarios that potential users deem challenging for AVs. (2) Implementation of a website[1] providing the capability to upload and annotate videos from the automotive domain to educate the general public. (3) Results of a preliminary study to confirm the findings by Colley et al. (2022) that participants do request more information when none is visualized (in the baseline). This shows that participants have an information deficit, which could lead to over- or undertrust. However, this visualized information does not significantly alter trust. (4) Results of a second study in which participants could upload their own videos. In these videos, the functionalities were then visualized automatically. We found that situation awareness increased significantly with the visualized AV functionalities. Measured with the Situation Awareness Rating Technique, situation awareness is a qualitative measure of the information provided and one's own understanding of the situation. Therefore, high situation awareness shows that more information was presented to users. However, trust was again not altered significantly. Furthermore, we found that most scenarios deemed potentially challenging for AVs included vulnerable road users such as pedestrians or pedalcyclists.

2. Related Work

We first introduce trust as a core concept for the introduction of AVs into general traffic. Besides, this work builds on related work on visualizations of AV functionalities. Furthermore, previous work in visualizing uncertainty in these functionalities informed our study. Finally, as we aim to provide the general public with information about AV functionalities, we report perceptions of AVs and how these were studied.

[1] The experiment A website is accessible at `https://visualizing-vehicle-ai-one.onrender.com/`; The experiment B website without the ability to upload videos is accessible at `https://visualizing-vehicle-ai-two.onrender.com/`; The source code is available at `https://anonymous.4open.science/r/VisualizingAutomatedVehicleFunctionalities-58E3`

2.1. Trust in Automated Vehicles

Based on the Extended Technology Acceptance Model by Ghazizadeh, Lee and Boyle (2012), trust influences Perceived Usefulness, Perceived Ease of Use, and Behavioral Intention to Use. These then influence the Actual System Use. Therefore, trust is a core concept for appropriately introducing AVs to the general public.

Lee and See define trust "as the attitude that an agent will help achieve an individual's goals in a situation characterized by uncertainty and vulnerability" (Lee and See, 2004, p. 51). A feedback loop with steps between automation and a user is present in their trust model (i.e., the passenger). Information on automation may be communicated in various ways, such as through a display. As the trustor (i.e., the passenger) assimilates this knowledge, the process of belief formation and trust evolution takes place. The trustor intends to rely on automation if a sufficient level of trust is attained. If so, the model suggests returning to the belief development phase and information assimilation. A trustee has three trust-building-relevant dimensions: Performance, Process, and Goal. We argue that for trust-building, AVs could "show the process and algorithms of the automation by revealing intermediate results" (Lee and See, 2004, p. 74) to make the automation understandable.

Körber (2019) based his trust definition on Lee and See (2004) and Mayer, Davis and Schoorman (1995). According to Körber, trust consists of the dimensions of Competence / Reliability, Understandability / Predictability, Intention of Developers, Familiarity, Trust in Automation, and Propensity to Trust. In this work, we used the questionnaire by Körber (2019) and focused on Predictability and Trust in Automation scales.

As AVs can not yet perfectly perform the driving task, calibrated trust should be the goal of communicating AV functionality. The term *Calibrated Trust* (Muir and Moray, 1996) refers to a state in which the user's confidence is commensurate with the functionalities of the automated system, therefore avoiding challenges associated with over- and undertrust. Thus, it is crucial to understand how users perceive the functionalities of the AVs.

Having outlined the foundational role of trust in AVs, we now explore various explanatory approaches that aim to enhance this trust through different types of information presentation.

2.2. Explanatory Approaches To Support Trust in Automated Vehicles

One of the first works to evaluate how to provide explanatory information to users in AVs was done by Koo, Kwac, Ju, Steinert, Leifer and Nass (2015). They showed that in AVs, providing explanatory information (answering the question *why* as opposed to providing behavioral (i.e., *how*) information) led to the highest trust.

In four driving scenarios, textual explanation types (no, simple, and attributional explanations) were also evaluated by Ha, Kim, Seo and Lee (2020). The simple and attributional explanations incorporated information about the *how* and *why* of specific actions. The attributional explanations included the attribution to the AV, thereby clearly establishing the link between the actions of the AV and the related information (e.g., simple: "Stopped after identifying the sudden appearance of a pedestrian in the road" (Ha et al., 2020, p. 276) versus attributional: "The autonomous vehicle stopped after identifying the sudden appearance of a pedestrian in the road" (Ha et al., 2020, p. 276)). Trust depended on the perceived risk (Ha et al., 2020). In high-risk situations, trust was highest with no explanation, while trust was highest in the low-risk situations with the attributional explanation. With the simple explanation, trust remained approximately the same for all levels of perceived risk. Omeiza, Kollnig, Web, Jirotka and Kunze (2021) showed that while causal explanations (i.e., explaining why something happened) enhance task performance (i.e., predicting the next AV action, determining road user actions, and general scenario assessment), they do not directly increase perceived trust in AVs. Finally, Woide, Colley, Damm and Baumann (2022) showed that transparent information about the system (i.e., the course and detected vehicles via a bird's eye view) must be verifiable for the user. Providing information that cannot be verified against the real world does not support trust-building. An example would be that the situation does not allow information about detected vehicles to be provided because no other vehicles are on the road.

This previous work, therefore, showed that explanations of AV behavior support trust. However, most of these approaches are textual and, thus, require a mental mapping between the information and the driving situation. Therefore, **situated visualizations** could provide an appropriate approach for communicating AV functionalities as less mental mapping might be required (Colley et al., 2020), which could reduce mental workload.

2.3. Visualized Functionalities of Automated Vehicles

Previous research examined various display types (e.g., Augmented Reality (AR) windshield displays (WSDs), LED strips, ambient light, and Head-Up Displays (HUDs)) for communicating a broad spectrum of information related to the driving task. This information can be broadly categorized into the functionalities of **Situation Detection**, **Situation Prediction**, and **Trajectory Planning**.

Regarding Situation Detection, Currano, Park, Moore, Lyons and Sirkin (2021) compared *no HUD* with a *minimalist* (showing the Situation Detection by highlighting pedestrians and showing detected signposts) and *complex* HUD design (minimalist plus speed, battery level, and an animation showing that the environment is scanned). Results depended on the scene's dynamics (i.e., roadway characteristics, behavior of the traffic participants, diversity of road users) and participants' reported driving style (i.e., No Style, Anxious, Dissociative, Impatient, Risky, Careful, and Stress Reduction). This resulted in the conclusion that an adaptive visualization in the HUD dependent on the user and the scene is necessary to increase the user perception of the AV (i.e., how users describe the AV). Colley, Krauss, Lanzer and Rukzio (2021b) also evaluated different visualizations of perceived objects by using a symbol combined with text in a HUD, an LED strip, or an AR WSD, either highlighting the object via a circle with text or by coloring in the detected object. They showed that no technologically challenging futuristic visualization technology capable of situated visualization (e.g., AR-based) is required but that current HUDs suffice. The ambient light was also used for visualizing the detection of, for example, pedestrians (Wilbrink, Schieben and Oehl, 2020).

Colley et al. (2020) visualized Situation Prediction (i.e., pedestrian intention) on a simulated AR WSD or a tablet. They found that the AR WSD led to lower cognitive load, potentially due to the situated visualization and, thus, reduced need for mapping information. Trust was also significantly higher with AR and visualization of three derived pedestrian intention levels (crossing the street, remaining on the sidewalk, uncertain) vs two states (certain and uncertain). Finally, in the baseline without visualization, the functionalities of the AV were rated significantly worse than with the visualizations, showing that these altered the user perceptions.

Regarding Trajectory Planning, Colley, Speidel, Strohbeck, Rixen, Belz and Rukzio (2024) visualized the AV's own trajectory besides the predicted trajectories of other vehicles (which represent Situation Prediction). Both trajectory visualizations incorporated uncertainty shown via color. They found that trust was significantly higher with than without the visualization of the own trajectory. However, no significant effects were found on the user perception of the AV functionalities. Similarly, Schneider, Hois, Rosenstein, Ghellal, Theofanou-Fülbier and Gerlicher (2021) evaluated explanations of the future trajectory via an AR WSD and an LED strip. First-time users reported increased user experience with explanations. The user experience could not be improved by combining this with a post-explanation via a smartphone app as a form of retrospective information. Also, perceived safety was not affected significantly. Trust or perceived functionalities were not evaluated. Besides the AR WSD, ambient light was used to inform the user about AV's Trajectory planning, such as starting or stopping (Wilbrink et al., 2020) and changing lanes (Löcken, Heuten and Boll, 2016). No empirical data on the effects of trust or other user perceptions are reported.

Some works also employed combinations of visualizing Situation Detection, Situation Prediction, and Trajectory Planning. Lindemann, Lee and Rigoll (2018) combined various information into one interface design. For example, they used AR WSDs to communicate detected objects (Situation Detection) that could lead to dangerous situations (e.g., pedestrians on the street) as well as derived behavior (Situation Prediction). Colored cubes visualized over vehicles indicated the derived behavior, e.g., the cube was green when the behavior was deemed safe and yellow for unusual or dangerous behavior. Using the interface results in higher trust. Colley et al. (2022) compared visualizing information on Situation Detection, Situation Detection, Trajectory Planning, and their combination. Regarding trust, they only found that showing the AV's planned trajectory led to reduced trust. Regarding user perceptions of functionalities, they found significant differences between no visualization and a range of combinations. Combining multiple functionalities generally led to reduced functionality assessments (e.g., pedestrian recognition was rated worse than in the baseline).

In summary, we found that previous work often combined multiple AV functionalities in its visualizations. However, numerous works do not include uncertainty information about the functionality. This prohibits the formation of calibrated trust, as the user will assume perfect functionality. Thus, in the next section, we report visualization approaches that include uncertainty in specific.

2.4. Automation Uncertainty Visualization

There have been few works that investigate the visualization of automation uncertainty. In this work, automation uncertainty represents the AV's fidelity to successfully accomplish its tasks (Situation Detection, Situation Prediction, or Trajectory Planning). These visualization approaches can be grouped based on their level of abstraction. Regarding high abstraction, Beller, Heesen and Vollrath (2013) showed a simple anthropomorphic symbol when system limits (i.e., high automation uncertainty) occurred, leading to increased trust and situation awareness. Helldin, Falkman, Riveiro and Davidsson (2013) used bars to indicate the operation certainty (they call it "reliability"). Contrary to Beller et al. (2013), who employed an emoji to display automation uncertainty, they found that this visualization reduced trust. With lower abstraction, Kunze, Summerskill, Marshall and Filtness (2018) used simulated AR WSDs to show uncertainties of longitudinal and lateral control, that is, the ability to steer and accelerate/decelerate. They compared different visual variables and found that the hue conveyed urgency the best due to automation uncertainty. Kunze, Summerskill, Marshall and Filtness (2019), arguing against the instrument cluster for uncertainty communication, employed a light strip as a peripheral cue and a vibrotactile seat. Peintner, Manger and Riener (2022) compared a baseline with a certainty (which they call "confidence") visualization regarding pedestrian crossings. A bar was compared to an alphanumeric display of the confidence in percent. They found that trust was lower and interaction took longer when communicating confidence value (bar or alphanumeric) compared to a baseline without the visualization.

Colley et al. (2021a) investigated the effects of visualizing the semantic segmentation task (regarding Situation Detection). They argue that previous studies used abstract representations that do not allow the user to identify the source of the uncertainty. They discovered that their simulated AR WSD did not increase trust or mental workload. However, users rated recognition-related attributes significantly higher. In a Wizard-of-Oz study in a real vehicle, Flohr, Valiyaveettil, Krüger and Wallach (2023) also found that this visualization increased understandability, perceived usefulness, and hedonic user experience. Furthermore, Kim, Yeo, Jo, Rus and Kim (2023) showed that this improved perceived usability, trust, and safety without adding cognitive load.

2.5. Investigating Public Perceptions of Automated Vehicles

Public encounters with AVs become more common, for example, in the form of Waymo's fleet in Phoenix or San Francisco[2]. However, this exposure also highlights widespread misconceptions about AV technology and functionality.

Prior research categorized public sentiment about AVs and identified the roots of these misconceptions. Du, Zhang, Liu, Xu and Liu (2022) categorized participants into four groups regarding their perception of AVs: "don't know" (19.2%), "neutral to positive" (32.6%), "naïve enthusiasts" (28.3%), and "sober skeptics" (19.9%). About a third believed AVs are already commercially available. Those with more misconceptions were more receptive to AVs, while better-informed participants were skeptical. They highlight the urgent need for "effective public communication to dispel myths about AVs and prevent AV technology from becoming controversial" (Du et al., 2022, p. 1). Besides, Cai, Jing, Wang, Jiang and Wang (2023) assessed the impact of "over-hype" on public misconceptions about AVs. They found that accident reports influenced misconceptions about AV availability (i.e., that AVs are already commonly available) but not safety. The most significant misconception arose from the overstatement of advanced driver assistance systems (ADAS) functionalities, especially regarding safety. Interestingly, younger and more educated individuals were more prone to these misconceptions. However, the effect of "over-hype" was consistent across diverse socio-economic groups (gender, age, income, education, career, driving license). Furthermore, Jing, Cai, Wang, Wang, Huang, Jiang and Yang (2023) analyzed public perceptions of AVs, especially after crash incidents, by drawing from 42,111 comments on Chinese social media platforms (Sina Weibo and TikTok). The findings indicate a tendency to blame vehicles with ADAS over human drivers after crashes and prevalent misconceptions (i.e., misinterpreting the functionalities) about AVs influenced by media and automotive promotions. These studies underscore the need for clear public communication on AVs to prevent misinformation and controversy.

Conclusion

The literature reviewed underscores the critical role of visualization technologies in enhancing user trust and understanding of AV functionalities. While different visualization technologies and visualization of functionalities have been evaluated, there is a missing understanding of (1) a more general population and (2), especially, which situations potential users find relevant for visualizations. So far, previous works have only evaluated the effect of the visualizations in scenarios pre-determined by the researchers. However, these also exhibit bias Innes and Fraser (1971), which could lead to scenarios being evaluated that do not reflect the experiences or situations relevant to the potential AV user. Besides, previous work showed public misconceptions lead to skepticism towards AVs. This requires improved communication of AVs' functionalities to calibrate user perceptions. Therefore, we designed and implemented a website addressing these shortcomings.

3. Implementation

In the following, we describe our visualization concepts for the driving scene videos used in the experiments. Besides, we explain our website architecture, which enables the automatic application of the visualization concepts on videos uploaded by the public and thus the application of the EDULICIT method.

[2] `https://waymo.com/waymo-one/`; Accessed on: 29.02.2024

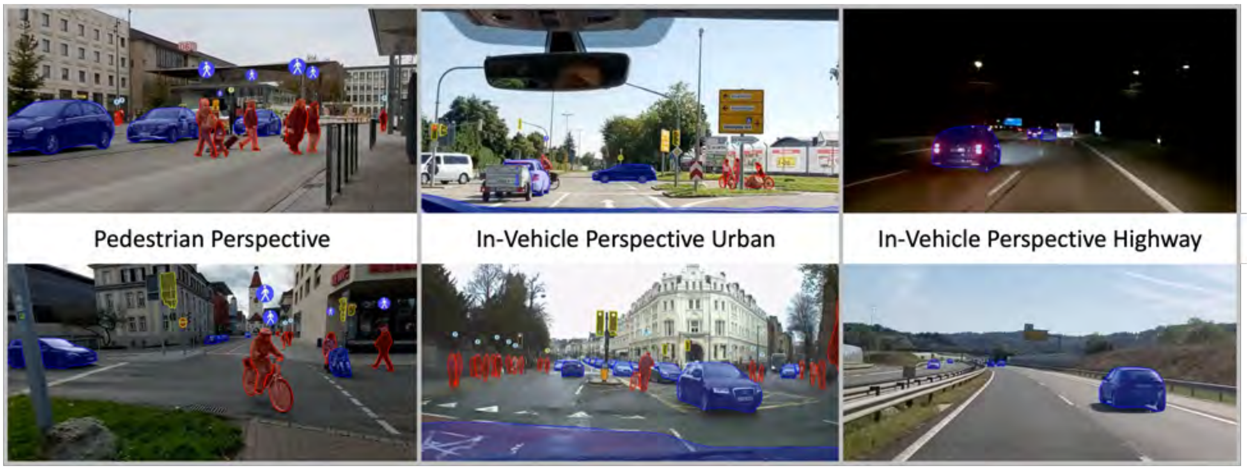


Figure 1: Screenshots of visualizations of automated vehicles' Situation Detection and Prediction in user-uploaded videos. The user-chosen scenarios range from pedestrian to in-vehicle perspectives of urban and highway traffic.

The visualization concepts used vary between the experiments due to technological constraints. We describe this more closely per functionality.

3.1. The EduLicit Method

In various fields, it is important to understand which use cases and hurdles exist for future technology adoption. However, this task is difficult, especially with novel, not yet-available, or fast-developing technologies. Asking potential users directly leads to relevant and specific information but is time-consuming (setting up the stimuli, survey, recruiting participants, evaluating attention checks) and can evoke biases, while existing opinions on future technologies, like data from forum entries, can be unspecific and difficult to analyze.

With EduLicit, we want to achieve two goals regarding the visualization of AV functionalities: (1) We want to *elicit* which scenarios potential users of AVs deem challenging for AVs. Eliciting scenarios requires the potential users to think of scenarios or to document scenarios they encounter in the real world (e.g., a busy intersection). Documenting these encounters helps to ground these scenarios in actual behavior. There are numerous options to collect this data. For example, this could be done via interviews, focus groups using speculative design (Galloway and Caudwell, 2018) or design fiction (Hales, 2013), descriptions, or sketches. Another possibility is to take videos, for example, with smartphones. (2) We want to provide a platform to *educate* potential users of current possibilities regarding the functionalities of AVs. Education is a complex topic that involves numerous possibilities for intervention. According to the cognitive-affective model of immersive learning (CAMIL) by Makransky and Petersen (2021), these factors include technological factors such as immersion but also affective and cognitive factors such as interest or motivation.

This work focuses on eliciting scenarios with EduLicit that potential users filmed in the real world. These videos provide rich detail of the scenario undistorted by the cognitive processes of the user (i.e., having to remember the scenario). Furthermore, these videos directly provide stimuli for the education part, as these can be used for the visualization of AV functionalities. Additionally, these videos interest the uploaders, thereby improving potential learning outcomes according to the CAMIL model. EduLicit, thus, stands for "Educating & Eliciting".

EduLicit is not intended to be restricted by technology and can be used with any widespread technology to educate and elicit relevant feedback and scenarios. For example, we employed a website in this work. This enables users to view the visualizations from their homes, allowing us to reach the majority of people. The website is accessible worldwide and can be quickly translated as only the texts need to be changed. Otherwise, even Google Translate in Chrome could already suffice for the translation. This can be used to reach more diverse populations. In the current implementation, we can educate about Situation Detection and Situation Prediction. By providing additional information, the website could also be leveraged to add more knowledge, thereby further improving education possibilities.

While EduLicit improves specificity about user data, guaranteeing relevant data by participants is impossible (e.g., the submitted videos could be unrelated to the traffic context). Nonetheless, even partially unrelated submissions can be interesting depending on the area of interest. For example, while our study was about visualizations in AVs, the actual videos of users can be from the perspective of a pedestrian and still highlight interesting scenarios.

By simultaneously focusing on the two aspects of *elicitation* and *education*, we believe that researchers can better fulfill their responsibilities to report interesting insights and serve the broader public.

3.2. Visualization Concepts

Analogous to Colley et al. (2022), we investigate the effects of visualizations regarding the functional implementation levels (1) *Situation Detection*, (2) *Situation Prediction*, and (3) *Trajectory Planning* of AVs (Dietmayer, 2016; Kunz, Nuss, Wiest, Deusch, Reuter, Gritschneder, Scheel, Stübler, Bach, Hatzelmann, Wild and Dietmayer, 2015; Taş, Kuhnt, Zöllner and Stiller, 2016). We propose visualizing the functionality in AR WSDs embedded in all windows (i.e., windshield and peripheral or side windows), which is preferable to tablet-based alternatives due to their situated visualization (Colley et al., 2020, 2021a) and a feasible output location (Jansen, Colley and Rukzio, 2022). In line with Colley et al. (2022), we assume that such visualizations calibrate trust as users are unfamiliar with the multitude of sensors in an AV's front, rear, and sides. We argue against using less distracting visualization methods, such as light bands (e.g., Wilbrink et al. (2020)), which are not feasible when multiple objects (e.g., pedestrians) overlap from the AV's point of view. Besides, such less-distracting methods would highlight fewer objects, from which users may assume that the AV did not detect them. Therefore, we suggest a more fine-grained visualization of objects relevant to the driving task.

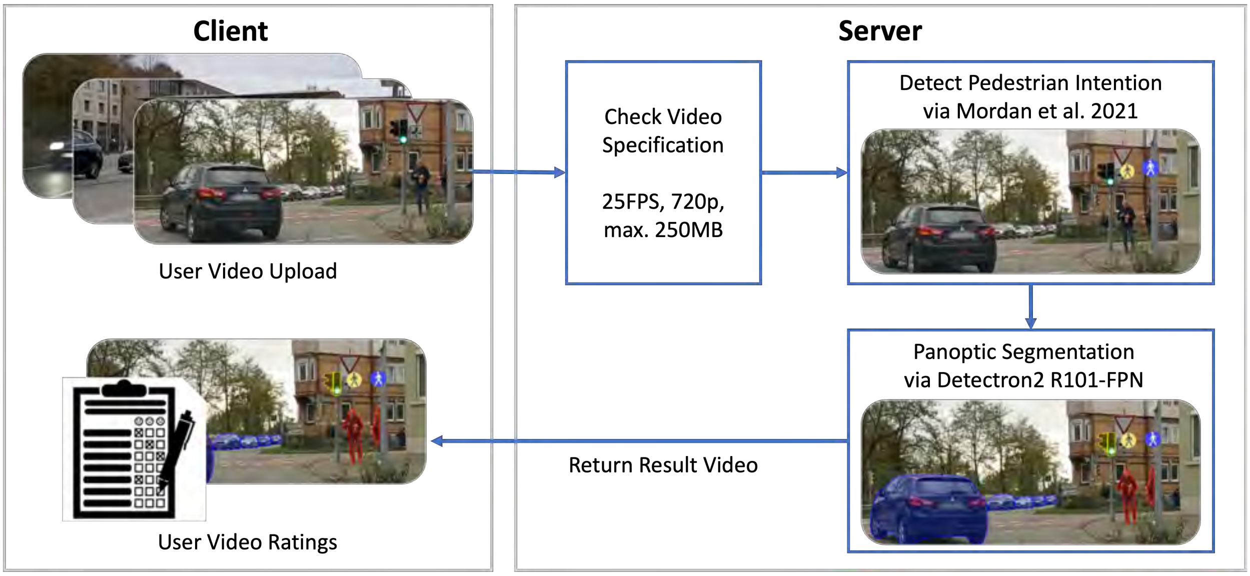


Figure 2: Overview of our video processing pipeline consisting of client and server. The client side enables users to upload their videos and rate the converted videos. The server processes the videos by detecting pedestrian intention via Mordan et al. (2021) and subsequently applying panoptic segmentation using Detectron2 R101-FPN (Wu et al., 2019) for Situation Detection.

Situation Detection On this level, AVs detect all objects in a situation (driving environment) using sensor data (e.g., via Lidar, GPS, and cameras) (Dietmayer, 2016). Therefore, in line with Colley et al. (2022), we visualized detected pedestrians and cyclists (both red), other vehicles (blue), and signposts (yellow) as they have a significant impact on the subsequent AV trajectory. In this case, semantic segmentation can encode uncertainty information as only detected objects are visualized (Colley et al., 2021a) (see Figure 1 and Figure 5). We used the panoptic segmentation of the Detectron2 R101-FPN by Wu et al. (2019).

Situation Prediction AVs predict the next situation state (e.g., next positions of vehicles and pedestrians) based on the preceding Situation Detection (Dietmayer, 2016). To estimate pedestrians' and bicyclists' intentions to cross the street, we employed the state-of-the-art pedestrian attribute recognition by Mordan et al. (2021). Similar to Colley et al. (2022), we utilized color-coded circles above the people's heads to visualize their intention to cross. A yellow circle indicates that the AV did not clearly recognize the intention. Light blue visualizes the intention to stay on the sidewalk, and dark blue visualizes the intention to cross the street (see Figure 1 and Figure 5). In contrast to Colley et al. (2022), we also apply our visualization concept to videos uploaded by the public (e.g., study participants). Automatic processing must be used for these videos. Therefore, we cannot inlay the vehicle intentions and trajectory prediction, for example, using a video editing program. To the best of our knowledge, there is currently no publicly available monocular vision-based approach to the automated prediction of other vehicles' trajectories. Consequently, we followed different approaches for each video variant. For the videos uploaded by the public (1), we decided against visualizing other's vehicle intentions. However, for the pre-defined videos (2), we used the same Situation Prediction proposed by Colley et al. (2022). In these videos, trajectories were visualized as successive arrows with a color gradient ranging from blue (certain) to pink (uncertain) (see Figure 4).

Ego-Vehicle Trajectory Planning AVs plan their next maneuver along a driving trajectory informed by the position of other road users and their predicted intentions (Dietmayer, 2016). We used the same visualization for the ego trajectory as for the trajectory prediction of other vehicles as Colley et al. (2022) (see Figure 4). However, as there are currently no publicly available monocular vision-based approaches to predicting ego-vehicle trajectories, we again cannot inlay the ego-vehicle maneuver planning in the videos uploaded by the public. Therefore, we only visualized the ego-vehicle maneuver planning in the pre-defined video variant.

3.3. Website Architecture

With our website (see Figure 3), we (1) hosted pre-defined videos provided by Colley et al. (2022) to compare our results to previous work and (2) enabled the general public to upload videos of driving situations they deem difficult for AVs. The website follows a basic client-server architecture. We utilized an Apache 2 web server on a Windows Subsystem for Linux (WSL2) configured as Ubuntu 20.04 LTS. The Windows host machine provides an NVIDIA RTX 3080 to apply the vision-based neural networks (see Section 3.2) on the uploaded video. On the client side, we created an HMTL-based website utilizing the Bootstrap 5 library for site layout and PHP for communication with the server.

Our pipeline for the **automatic** conversion and return of videos uploaded by the public consists of five steps (see Figure 2): (1) First, website visitors (e.g., study participants) see the landing page and receive an overview of the features and instructions on the video upload (see Figure 3 a). (2) After reading the instructions, they upload their videos, which are sent in chunks to the server to minimize the conversion complexity later in the pipeline. (3) On the server side, a Python script automatically rejects videos exceeding the allowed specifications: a maximum of 25 FPS, a resolution of 720p, and a maximum file size of 250 MB. The specifications were chosen to reduce the video conversion runtime to not overload the server in case of many concurrent visitors. (4) If the video meets the specifications, the server starts the vision-based neural networks to apply pedestrian intention and panoptic segmentation to the uploaded video.

The visualization models run **sequentially** on the video to generate a single result video. To reduce possible interference (e.g., due to bold segmentation borders), we decided to apply the pedestrian intention model (Mordan et al., 2021) first, which makes fewer visual changes to the source video than the panoptic segmentation model (Wu et al., 2019) used for Situation Detection. The panoptic segmentation model then runs on the resulting video of the pedestrian's intention. This application order returned promising results in internal tests with videos of arbitrary visual complexity (i.e., varying numbers of scene objects, such as vehicles and pedestrians). The resulting videos in these tests showed no obvious visual differences compared to videos from applying a single model. Therefore, we argue that the proposed model application order has a negligible impact on the resulting video.

(5) Finally, the video with the visualized Situation Detection and Prediction is shown to the website visitor (see Figure 3 d). Visitors can then answer questionnaires embedded into the website regarding the converted version of their uploaded videos (see Figure 3 e).

By allowing easy exchange of visualizations and/or neural networks, our website helps researchers to elicit and educate the public in numerous ways, thereby allowing a more widespread application of EduLicit. A limitation of this approach lies in the ambiguity between the performance of openly available networks compared to commercially available software. As major commercial vendors such as Mercedes-Benz AG, Tesla, or Waymo do not provide sufficient information about their performance, comparing the website's results with commercial implementations is impossible.

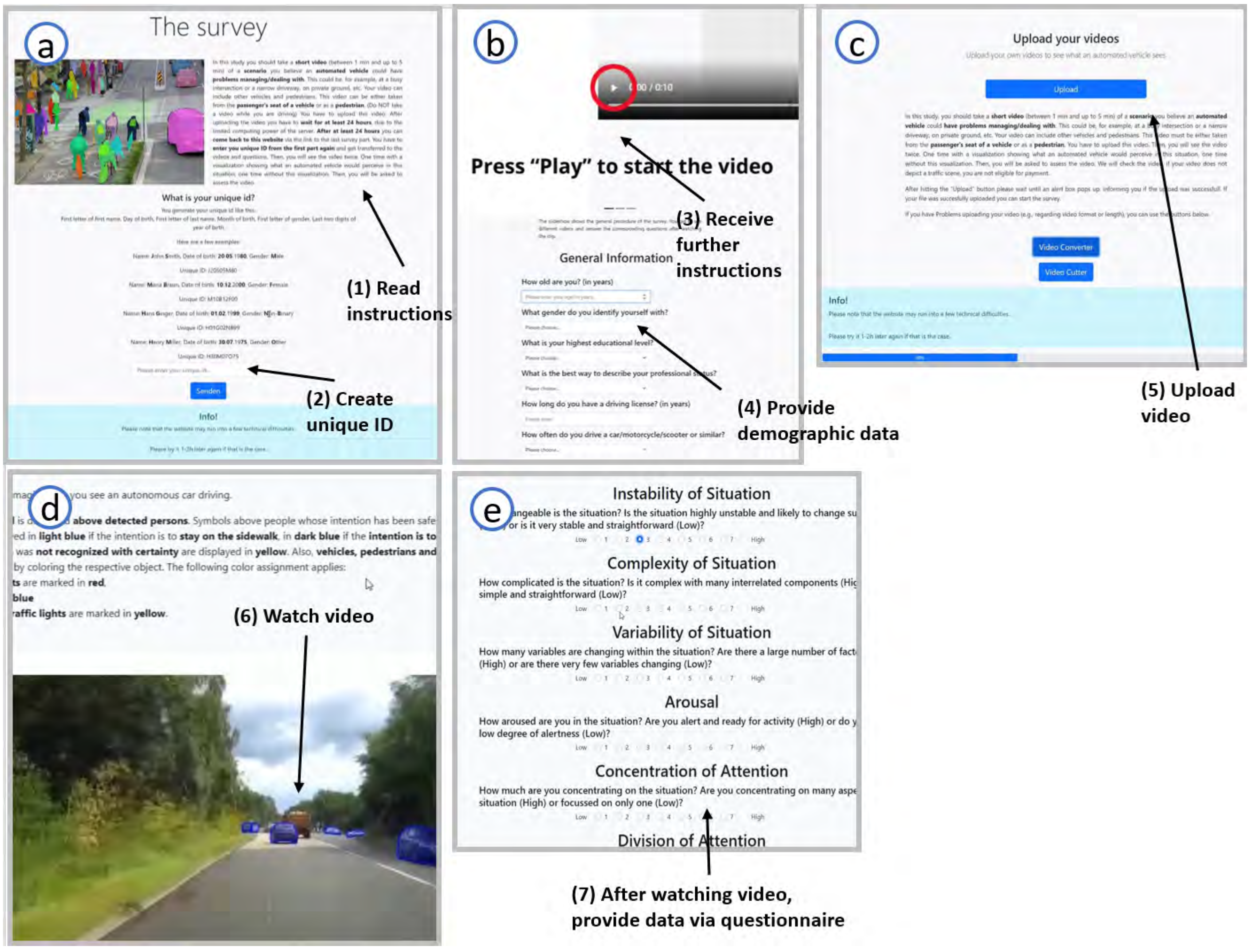


Figure 3: Screenshots of our website enabling users to upload self-chosen scenarios and receive visualizations of AVs' Situation Detection and Prediction. **(a)** Landing page with survey instructions. **(b)** Providing demographic data. **(c)** Uploading a video with utility websites linked. **(d)** Video returned by our visualization pipeline (see Figure 2). **(c)** Questionnaire page to retrieve user ratings. The screenshots were cropped for better visibility.

4. Experiment A: Replicating Previous Approaches

4.1. Research Goal

We designed and conducted a within-subject study to replicate a subset of the measured effects of AV feedback visualizations by Colley et al. (2022). Therefore, this study was guided by the research questions:

RQ1 *What impact does the visualization of Situation Detection, prediction, and ego maneuver planning have on passengers in an AV in terms of (1) cognitive load, (2) trust, (3) perceived safety, (4) preference, (5) subjective situation awareness, and (6) capability assessment?*

RQ2 *Are the results from the previous work by Colley et al. (2022) replicable?*

Participants experienced two conditions: a baseline without and a video with the visualizations. Thereby, we replicate the experiment to confirm the reliability and validity of the findings, ensuring they are not artifacts of specific conditions or chance. This aids in identifying errors, enhances generalizability across different contexts, and spurs innovation by uncovering new research avenues when discrepancies arise. This replication also highlights the appropriateness of using the website.

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

Figure 4: Screenshots of the two videos used in *experiment A* with permission from Colley et al. (2022). **(a)** Baseline with no visualization. **(b)** Combined visualization of Situation Detection, prediction, and maneuver planning.

4.2. Materials

For the first experiment, we used two videos provided by Colley et al. (2022). We used the baseline without any visualization and the video with object detection, prediction, and ego maneuver visualizations, that is, the video with **all** visualizations.

4.3. Procedure

The session started with an introduction, agreeing to the consent form, and a demographic questionnaire. Both videos were shown in random order. Following the condition, participants completed the questions outlined below. Finally, participants were invited to provide general feedback. A session lasted 12 minutes on average. Participants received £1.20 in compensation.

4.4. Measurements

We used the same measurements as proposed by Colley et al. (2022). Thus, we measured cognitive load with the mental workload subscale of the raw NASA-TLX (Hart and Staveland, 1988) on a 20-point scale ("How much mental and perceptual activity was required? Was the task easy or demanding, simple or complex?"; 1=*Very Low* to 20=*Very High*). We measured situation awareness using the situation awareness rating technique (SART) (Taylor, 2017). As done by Colley et al. (2022), we used the subscales *Predictability/Understandability* (*Understandability*) and *Trust* of the *Trust in Automation* questionnaire by Körber (2019). Participants also stated perceived safety using four 7-point semantic differentials from -3 (anxious/agitated/unsafe/timid) to +3 (relaxed/calm/safe/confident) (Faas, Kao and Baumann, 2020). We used the scales for usability proposed by Colley et al. (2022) and closely related to the System usability Scale Brooke et al. (1996): "I think that I would like to use these visualizations frequently." and "I found the visualizations unnecessarily complex." (1=*Strongly Disagree* to 5=*Strongly Agree*). We also used the subscales *Performance*, *Judgement*, and *Reaction* of the Situational Trust Scale for Automated Driving (Holthausen, Wintersberger, Walker and Riener, 2020).

We also let participants rate AVs' functionalities. These were assessed using self-developed single items from Colley et al. (2022): driving style (1=completely safe to 7=completely dangerous), object detection (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", and lateral and longitudinal guidance on 7-point Likert scales (1=*Totally Disagree* to 7=*Totally Agree*).

4.5. Results of Replication Experiment

Data Analysis We used R 4.3.3 and RStudio 2023.12.1. All packages were up-to-date in April 2024. For post-hoc tests, we used Bonferroni correction. We use Wilcoxon's test (in case of non-normally distributed data) or Student's t-test to account for the within-subject study design and compare the conditions.

Participants We computed the desired sample size before the experiment via an a-priori power analysis using G*Power (Faul, Erdfelder, Buchner and Lang, 2009). To achieve a power of .8, with an alpha level of .05, 54 participants

Table 1
Statistical information about experiment A. For the explanation of the ✔and ?, see Section 4.7. Sig.=Significance; RR=Replicated Result; Rec.=Recognition

Dep. Variable	Statistic	Sig.	Effect Size	Baseline	Visualization	RR
Cognitive Load	V=825.00, p=0.004	**	r=-0.39	**M=9.54**	M=12.03	?
Situation Awareness	t(75)=-1.61, p=0.11			M=20.55	M=22.49	✔
Predictability	V=1120.50, p=0.37			M=3.39	M=3.56	✔
Trust	V=1025.50, p=0.48			M=3.07	M=3.22	✔
Perceived Safety	V=1174.00, p=1.00			M=0.96	M=0.90	?
Rec. of pedestrians	V=972.50, p=0.51			M=5.01	M=4.83	✔
Rec. of signposts	V=938.50, p=0.86			M=5.04	M=5.00	✔
Rec. of pedestrian intention	V=1145.00, p=0.35			M=4.83	M=4.58	✔
Rec. of vehicle paths	V=743.50, p=0.83			M=5.09	M=5.13	✔
Longitudinal Control	V=983.50, p=0.79			M=5.24	M=5.21	✔
Lateral Control	V=647.50, p=0.54			M=5.47	M=5.59	?
Performance	V=983.50, p=0.33			M=3.80	M=4.13	✔
Judgment	V=692.50, p=0.84			M=2.09	M=2.18	✔
Reaction	V=692.50, p=0.84			M=5.68	M=5.72	✔
Reasonability	V=902.00, p=0.47			M=4.97	M=5.20	✔
Necessity	V=902.00, p=0.47			M=4.96	M=5.05	✔
Visual Clutter	V=132.50, p<0.001	***	r=-0.85	**M=1.95**	M=3.99	✔
Perception-related info	V=1342.00, p<0.001	***	r=0.62	M=5.20	**M=3.92**	?
Prediction-related info	V=1148.00, p=0.02	*	r=0.34	M=5.14	**M=4.46**	✔
Maneuver-related info	V=1167.00, p=0.02	*	r=0.32	M=5.17	**M=4.53**	✔

should result in medium effect size (*Effect Size dz*=0.5; based on (Colley et al., 2021a)) in a two-tail t-test with matched pairs.

We recruited N=100 participants. 24 had to be excluded due to technical issues. Therefore, we had a final sample of N=76 (38 female, 36 male, 2 non-binary) from `prolific.co`. To prevent confounding effects of traffic handedness (right-hand vs. left-hand traffic) or culture, the participant pool was limited to US residents (Rasouli and Tsotsos, 2019). Moreover, we targeted US residents to align the sample with North America[3] as one of the primary markets for AVs.

Participants indicated that their highest educational level was College (52), followed by High School (19), Secondary School (1), and Vocational Training (4). Regarding their employment status, 35 are employees, 7 are students at a college, three are at a school, 16 are self-employed, five are job-seeking, and 13 indicated *other*. This range from high school to college and from job search to study and self-employment shows that users from diverse educational and employment backgrounds participated in the study, corresponding to the potential future target audiences of AVs across society (Kyriakidis, Happee and de Winter, 2015).

On average, participants were M=38.28 (SD=14.23) years old. All participants hold a valid driver's license for, on average, M=18.84 (SD=14.99) years. This indicates that the participants have extensive driving experience relevant to understanding traffic problems, as they have probably experienced difficult scenarios themselves from a driver's perspective. On 5-point Likert scales (*1=Strongly Disagree — 5=Strongly Agree*), participants showed medium interest in AVs (M=3.33, SD=1.36), believed AVs to ease their lives (M=3.58, SD=1.22), and believed AVs to become a reality by 2032 (M=3.66, SD=1.09).

All Results The statistical analysis results can be seen in Table 1. These results answer RQ1 by showing the impact of the visualizations on the dependent variables.

[3] `https://www.thebrainyinsights.com/report/autonomous-vehicle-market-13887`; Accessed: 29.02.2024

4.6. Open Feedback

We received mostly positive feedback regarding the visualizations. Two participants were confused by the flickering of the visualization (*"a view of the information WITHOUT the flickering could be much more informative, the flickering makes it tough to understand what is what"; "I thought the second video with all the frippery, bells and whistles was so distracting AND made a 2000+ pound car / lethal weapon look waaayyy too much like a 2d game"*) despite it showing the uncertainty. One participant even felt that the visualization was too stressful. However, it made the person cautious (*"wow that was stressful. I just don't know how anyone trained to drive for decades could get used to even riding along in a car that is guided this way. I felt like it was going to hit someone constantly. people are unpredictable. Even if it predicts their movements like, it can't really be 100% with those predictions. I also thought it felt like it was driving awfully fast. "*). While not every participant provided feedback, others were positive. For example, participants mentioned that they trusted the AV more and that it was better to know what was happening (*"The first video was great to see all the information about how the vehicle was interpreting the situation and acting accordingly. I felt like I could trust the car a lot more."; "The visualizations from the car were a bit much to take in all at once, but knowing what was going on "in the car's head" was less anxiety-inducing than not knowing"*).

4.7. Comparison with Previous Work & Conclusion

We compared the videos representing conditions 1 (no visualization) and 8 (showing Situation Detection, Prediction, and Trajectory Planning) by Colley et al. (2022). Therefore, we did not completely replicate their work. We compare our results regarding significance with their work per dependent variable. We determined that our approach could replicate the results if the same tendency were found (indicated via a ✔; see Table 1). If there was a significant test result but no significant post-hoc test, we determined that it is unsure whether we could replicate the results (indicated via a ?; see Table 1). We found that we could replicate almost all results, thus showing that the effects found by Colley et al. (2022) are valid and that our tool can replicate these. Therefore, we can clearly state that RQ2 can be affirmed. This experiment used a pre-recorded video to determine how uncertainty visualizations affect potential users. It used a scenario with multiple pedestrian crossings, with pedestrians of varying ages. However, this only represents one scenario of a large potential space of uncertain situations. As we wanted to know which situations would be perceived as challenging, we conducted a second experiment, allowing participants to upload their videos.

5. Experiment B: Videos Uploaded by the Public

5.1. Research Goal

After replicating the results of Colley et al. (2022), we were interested in probing the public about the situations in which they imagine AVs to face challenges. Additionally, we were interested in whether visualizations could aid here too. Therefore, our research questions were:

RQ3 *Which specific scenarios do people find challenging regarding AVs in everyday traffic?*

RQ4 *What impact does the visualization of Situation Detection and Prediction of pedestrians have on passengers in an AV in terms of (1) cognitive load, (2) trust, (3) perceived safety, (4) preference, (5) subjective situation awareness, and (6) capability assessment for their own uploaded scenarios?*

We only included the visualization of Situation Detection and Prediction of pedestrians as these are feasible via the open source models of Mordan et al. (2021) and Wu et al. (2019). A manual prediction of other vehicles' trajectories or the ego maneuver as done by Colley et al. (2022) was not feasible (see Section 3.2).

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

5.2. Materials & Procedure

In this experiment, participants were asked to "take a short video (between 1 minute and up to 5 minutes) of a scenario you believe an automated vehicle could have problems managing/dealing with. This could be, for example, at a busy intersection or a narrow driveway, on private ground, etc. Your video can include other vehicles and pedestrians. This video can be taken from the passenger's seat or a pedestrian's point of view. (Do NOT take a video while you are driving)". The experiment ran, in total, for six weeks. Participants were able to upload their videos from their personal computers. After the server processed the uploaded videos, participants were assigned in counterbalanced order to the video processing conditions (Baseline or with visualization). The participants then rated their processed self-uploaded videos. As the processing depended on the video duration (e.g., a 5-minute video required approximately one hour of processing), participants were instructed to watch the videos the following day to ensure the processing was finished. All participants viewed the processed and unprocessed video (constituting a within-subject design). Each participant could upload one video and only rate their video. As we do not publicly show the recordings, no consent is necessary. Additionally, §23 KunstUrhG would allow to publicly show public areas (such as roads) if portrayed people are not the main focus (i.e., called "Beiwerk" in German or accessory).

5.3. Measurements

We employed the same measurements as described in Section 4.4. We also included a qualitative analysis of the uploaded videos (see Section 6.0.2).

5.4. Results of Experiment B

5.4.1. Data Analysis

We employed the same data analysis as in experiment A (see Section 4.5).

5.4.2. Participants

We again computed the desired sample size before the experiment via an a-priori power analysis using G*Power (Faul et al., 2009). To achieve a power of .8, with an alpha level of .05, 20 participants should result in medium effect size (*Effect Size dz*=0.6) in a one-tail Wilcoxon signed-rank test with matched pairs. Additionally, due to the included qualitative focus (see Section 6.0.2), a lower number of participants was deemed satisfactory while still being comparable to previous work (e.g., Colley et al. (2021b)).

We recruited 22 participants through flyer advertisements in our university's buildings (Mean age=29.0, SD=15.6, range: [18, 81]; Gender: 36.4% women, 63.6% men, 0.00% non-binary; Education: College, 50.00%; High school, 40.91%; Middle school, 4.55%; Secondary school, 4.55%). Participants in experiment B did not participate in experiment A. 14 participants are students at the college, two at a school, five are employees, and one declared "other". All participants hold a valid driver's license for, on average, M=10.00 (SD=11.52) years. On 5-point Likert scales (*1=Strongly Disagree — 5=Strongly Agree*), participants showed medium to high interest in AVs (M=4.18, SD=0.96), believed AVs to ease their lives (M=3.86, SD=1.17), and did not believe AVs to become a reality by 2032 (M=2.45, SD=1.30). Participation in experiment B was voluntary.

This sample captures the perspective of a younger, technologically adept demographic likely to be early adopters of AV technology (Kyriakidis et al., 2015). The educational backgrounds, which is a relevant factor for understanding technology adoption among young adults (Czaja, Charness, Fisk, Hertzog, Nair, Rogers and Sharit, 2006), were diverse and included college and high school.

6. Results

The statistical analysis results can be seen in Table 2. These results answer RQ4 by showing the impact of the visualizations on the dependent variables.

6.0.1. Open Feedback

Participants were surprised by the functionalities of the simulated AV, both positively and negatively. This is directly perceivable in the comments: "Interesting that the car sees pedestrians in a window of a shop and that some cars are not detected if they are driving right in front of the car."; "Some sign post where detected as pedestrians even though they are only 50 − 100m away and some cars aren[']t detected at all. The street was also su[r]prisingly often detected as a car."; "Some pictures of people on a bus driving by were detected as pedestrians. And some trash containers as cars. And most of the actual pedestrians in the scene were detected as wanting to cross the street but for me it seemed that most where just waiting."; "Almost every pedestrian intention who was walking on the sidewalk was detected as wanting to cross the street even though they are just walking there."; "Some construction posts are detected as pedestrians (−) If the car is close enough the driver is also detected as a pedestrian (+) I personally was a few times dist[r]acted by the all the flashing and blinking and perceived some cars or other pedestrians a little bit later (−)"; "Some cars are not detected and some things far away are detected as people. And not all signs are detected."; "Some big trucks are not detected as a vehicle and almost every post on the edge of the road was detected as a pedestrian before the car was moving closer to it."; "I find it shocking how many "false" pedestrians the car detects and even some "false" cars too.";

Table 2
Statistical information about experiment B.

Dep. Variable	Statistic	Sig.	Effect Size	Baseline	Visualization
Cognitive Load	$V=48.50$, $p=0.11$			$M=11.82$	$M=13.14$
Situation Awareness	$t(21)=-3.65$, $p=0.001$	***	$r=-0.75$	$M=16.36$	**M=20.32**
Predictability	$t(21)=-1.68$, $p=0.11$			$M=3.02$	$M=3.30$
Trust	$V=101.00$, $p=0.82$			$M=3.11$	$M=3.09$
Perceived Safety	$V=109.50$, $p=0.31$			$M=1.51$	$M=1.23$
Recognition of pedestrians	$V=95.00$, $p=0.72$			$M=4.73$	$M=4.73$
Recognition of signposts	$V=133.50$, $p=0.007$	**	$r=0.75$	$M=5.14$	$M=3.68$
Recognition of pedestrian intention	$t(21)=1.94$, $p=0.07$		$M=4.77$	$M=3.86$	
Recognition of vehicle paths	$V=68.00$, $p=0.70$			$M=4.95$	$M=5.00$
Longitudinal Control	$V=22.50$, $p=0.18$			$M=5.32$	$M=5.59$
Lateral Control	$V=10.00$, $p=0.02$	*	$r=-0.74$	$M=5.23$	**M=5.73**
Performance	$V=86.00$, $p=0.03$	*	$r=0.64$)	$M=3.64$	$M=3.09$
Judgment	$V=29.50$, $p=0.43$			$M=2.64$	$M=2.45$
Reaction	$V=21.00$, $p=0.15$			$M=4.82$	$M=5.14$
Reasonability	$V=0.00$, $p<0.001$	***	$r=-1.00$	$M=1.73$	**M=5.50**
Necessity	$V=3.00$, $p<0.001$	***	$r=-0.97$	$M=1.86$	**M=4.91**
Visual Clutter	$V=13.00$, $p=0.001$	***	$r=-0.85$	**M=1.27**	$M=2.18$
Perception-related information	$V=210.00$, $p<0.001$	***	$r=1.00$	$M=6.59$	**M=3.18**
Prediction-related information	$V=87.00$, $p=0.004$	**	$r=0.34$	$M=6.55$	**M=5.55**
Maneuver-related information	$V=68.00$, $p=0.02$	*	$r=0.74$	$M=6.55$	**M=6.05**

"blue signs on the highway were not recognized in the video(no visual yellow marking) and the trailer behind a car was not recognized, only the car was marked in blue once it was visible"; "some signs were not marked for example on that specified the maximal speed of 60 kmh, also a trailer behind a car was not marked, only the car itself was marked in blue"; "Not all signs and vehicles are recognized"; "car coming from the right at the beginning was not recognized very fast, if it hasn't stopped there may have been a crash".

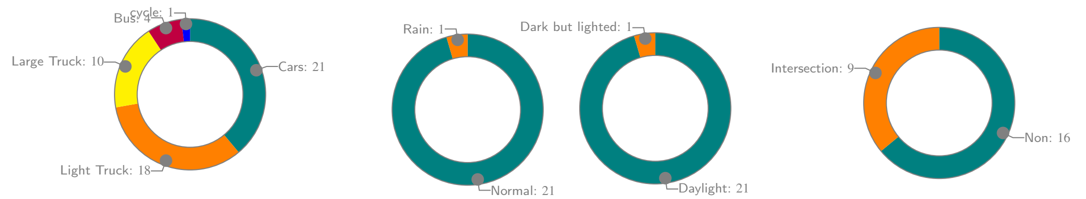
Figure 5: Exemplary screenshots of user-uploaded videos of driving scenarios they deem challenging for AVs' Situation Detection and Prediction. The user-chosen scenarios varied in criticality and complexity: **(a)** Mixed traffic between pedestrians and vehicles in a construction area. **(b)** Urban traffic at dusk. **(c)** Single-lane road with an adjacent pedestrian zone. The annotated video variants resulting from our visualization pipeline (see Figure 2) are depicted in the bottom row.

6.0.2. Qualitative Analysis of the Uploaded Videos

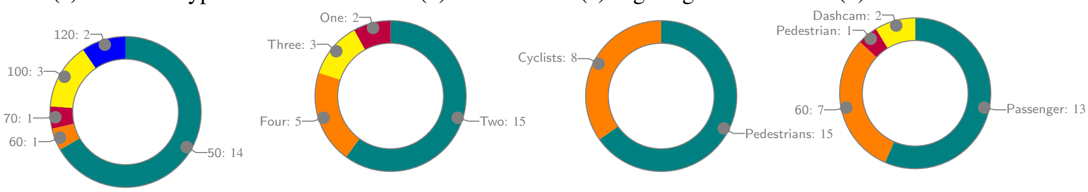
Two authors independently coded the videos. The first three videos were coded collectively to build a clear understanding of the coding. Afterward, both authors coded the remaining videos independently. One coding difference was resolved via discussions. The code book included **Weather** (Normal, Rain, Snow/Sleet, Other, Unknown), **Lighting** (Daylight, Dark but lighted, Dawn or Dusk, Other), **Present Vehicle Types** (Car, Light Truck (consisting of Pickup, Utility, and Van), Large Truck, Motorcycle, Bus), **Junction** (Nonjunction, Junction-Intersection, Junction-Intersection-Related, Other/Unknown), **Number of Lanes** (One, Two, Three, Four, More than four, Unknown), **Non-Occupants** (Pedestrians, Pedalcyclists, Other/Unknown), and **Speed limit**. These were taken from the NHSTA Traffic Safety Facts 2020 (Administration, 2023).

Additionally, we added the Viewpoint Video Taken (Dashcam, Passenger, Driver, Unknown, Pedestrian) and, if possible, the approximate location of the video. If a video contained more than one category, this was coded accordingly. This was done to answer our RQ3.

Results: Figure 6 shows the coding results. All self-taken videos were recorded near Ulm, Germany. Two videos were taken from online sources. Generally, the situation was mostly recorded during the day (see Figure 6c) in normal weather (see Figure 6b). All scenarios included various vehicle types (see Figure 6a). While the number of lanes varied (see Figure 6f), the multitude of videos contained pedestrians and/or pedalcyclists (see Figure 6g).



(a) Available types (b) Weather (c) Lighting (d) Junction



(e) Speed Limit (f) Number of Lanes (g) Non-occupants (h) Viewpoint

Figure 6: Video codings. Numbers represent the occurrences in the respective category. Passenger Cars=Cars, Pedalcyclists=cyclists, Motorcycle = cycle.

7. Discussion

In this work, we first presented the method EduLicit to simultaneously elicit scenarios deemed challenging for AVs while educating the public. To this end, we built a website allowing the general public to assess two neural networks relevant to Situation Detection (Wu et al., 2019) and Situation Prediction (Mordan et al., 2021), two main functionalities relevant to the driving task. In the first experiment (see Section 4), we replicated previous studies' results with N=76 participants regarding the effects of visualizing AV functionality with incorporated uncertainty. In the follow-up experiment (see Section 5), participants could upload videos of situations they deemed challenging for AVs. Our website then automatically visualized the AV functionality of Situation Detection and Situation Prediction using the results of the two neural networks for N=22 participants.

7.1. On the Elicited Scenarios

Our results show that the user-chosen videos primarily contain situations with many pedestrians. Vehicles were not the main focus. This implicit importance ranking is also reflected by the selection of videos showing single-lane roads and public places crowded with people, for example, near train stations and bus stops. We argue that this may hint that *humans come first, vehicles second*, as participants deemed situations with many pedestrians (and also bicyclists) as more complex for AVs than situations with (only) vehicles. This is also in line with participant feedback that

"...people are unpredictable..." (see Section 4.6) and their specific focus on pedestrian detection and intention prediction (e.g., "...shocking, how many false pedestrians the car detects...", see Section 6.0.1). Therefore, users could perceive the detection and prediction of humans as more difficult and, thus, unreliable. In this regard, the users' perception may be influenced by their own experiences. For example, when driving manually, pedestrians and cyclists are often challenging to see (especially in the dark), even to the human eye.

A more speculative interpretation would be that participants might not have considered that vehicles have human drivers, who are often also unpredictable. If participants did assume traffic without human drivers (i.e., future homogeneous AV traffic), this would change their assessment of a challenging situation for AVs. Besides, this may explain their focus shift on pedestrians instead of vehicles, as they could have assumed that not-connected traffic objects (e.g., pedestrians) are the remaining uncertainty factors in situations with interconnected AVs.

The viewpoint was mixed in many videos, with 34% pedestrian and 56% passenger points of view. We assume this distribution was mainly due to the difficulty of recording relevant driving scenarios. The participants had to be passengers and needed an additional driver for the video recording. Therefore, many participants opted for pedestrian point-of-view videos. However, participants may have recorded the pedestrian point of view on purpose. In this case, they consider the role of a pedestrian in traffic with AVs to be particularly difficult because they feel unsafe and mistrust the AVs. This highlights the importance of researching external communication between AVs and other traffic objects.

The selection of videos also highlights little diversity in the scenario context (weather conditions were mostly normal, and the time of day was mostly daylight). This could be due to higher convenience for video taking, or that context was perceived as a less important factor than traffic objects (e.g., vehicles, pedestrians, or bicyclists). However, weather conditions and lighting are among the most critical factors for safe AVs (see Vargas, Alsweiss, Toker, Razdan and Santos (2021)), especially for vision-based systems, as assumed in our experiments. Therefore, we argue that a website to educate the public should show a wide range of scenarios with regard to the factors named by Vargas et al. (2021). Nonetheless, our approach to visualizing AV's current Situation Detection and Situation Prediction could increase understanding.

In general, the user-chosen driving scenarios were not overly complex. We assume the study participants primarily wanted to see how the website converts their video. Therefore, some submitted a generic driving video recorded on a highway. However, they were still interested in AVs' functionalities of detecting and predicting driving situations. We argue that this is a first step towards achieving user acceptance, as people need a rough understanding of the technology (in this case, automation) before they can think about personal requirements for the technology. For example, which driving situation they deem difficult and whether pedestrians or vehicles are more challenging to predict.

7.2. Comparing Scenarios from Literature with User-Chosen Scenarios

Previous work has defined numerous scenarios where the AV should communicate functionalities to the user. Wiegand, Eiband, Haubelt and Hussmann (2020) identified 17 situations in which explaining communication could be necessary. 26 participants concluded that for unexpected driving behavior, there are six main concerns: emotion and evaluation, interpretation and reason, vehicle capability (which we called functionality), interaction, future driving prediction, and explanation request times (Wiegand et al., 2020). Some of these scenarios can be addressed by improving the automation alone, for example, an abrupt stop at a right turn, a long wait at an intersection to turn left, unnecessary lane change, strong brakes, quick decisions, and the car being very slow. However, others are only partially addressable by improved automation: Reluctant turn right due to a pedestrian (scenario 2 (Wiegand et al., 2020)), another car stopping, and a child crossing (scenario 9 (Wiegand et al., 2020)). Avetisyan, Ayoub and Zhou (2022) also based their work on the scenarios of Wiegand et al. (2020): waiting for a pedestrian, long wait at an intersection, AV stops for a pedestrian, stop due to emergency vehicle, strong braking to reach the speed limit, lane change due to heavy traffic, long wait before merging. Wintersberger, Nicklas, Martlbauer, Hammer and Riener (2020) identified six themes for information visualization: predictability, impact, object characteristics, spatial properties, regulations, and visibility. These were derived from assessing eight scenarios, including scenarios with differing numbers of lanes, traffic, and vulnerable road users (see (Holländer, Colley, Rukzio and Butz, 2021)). They concluded that there exists "an inverse correlation between situational trust and participants' desire for feedback" (Wintersberger et al., 2020, p.1) Colley et al. (2021b) considered three scenarios: one where a child crosses the street to run after a ball, a blind person crossing, and a dog crossing the road. Löcken, Frison, Fahn, Kreppold, Götz and Riener (2020) evaluated three scenarios: A pedestrian crossing the street, a car exiting a parking lot, and a bike crossing the street are all examples. Lanzer, Stoll, Colley and Baumann (2021) evaluated two context factors: Whether a pedestrian crossed with a smartphone in hand or without and the situation's complexity (modeled via the number of other road users).

Our results via the user-chosen scenarios mostly reflect these envisioned scenarios. However, the majority focused on either intersections or scenarios with pedestrians. While this could be due to the nature of the recording (participants most likely did not wait for a long time to experience specific scenarios), we also assume that this reflects their mental model that AVs will face challenges with large intersections and pedestrians. This also seems to resemble scenarios in which human drivers face difficulties. In the future, enabling more people to upload videos could elicit other scenarios.

7.3. Differing Results from Scenarios

In the replication experiment (see Section 4), we could assess the effects of visualizing the combination of Scenario Detection, Prediction, and ego Trajectory Planning in one scenario. In experiment B (see Section 5), the results were difficult to accumulate as the uploaded videos differed notably **and** the visualization only included Situation Detection and Situation Prediction for pedestrians (i.e., intention prediction). Nonetheless, we want to discuss the similarities and differences between the two experiments. We found similarities in the questions regarding the need for additional information. In both experiments, participants required additional information in the baseline condition, where no visualizations of AV functionalities were visible in the videos. In both experiments, participants felt the additional information conveyed by the visualizations caused visual clutter. However, in experiment A, this feeling was stronger. Cognitive load was not significantly different in experiment B. Therefore, we conclude that experiment A's additional visualization of trajectory planning was the "tipping point". While reasonability and necessity of visualizations were rated higher with the visualization in both experiments, this was only significant in experiment B. Also, only in experiment B was a significant improvement in visualization regarding perceived lateral control and situation awareness. Therefore, we conclude that in both experiments, the visualizations were deemed mostly positive but that future work is necessary to evaluate the underlying mechanisms.

7.4. Limitations from the Functionality Perspective

Our visualization approach uses monocular vision-based models for Situation Detection and Prediction. However, other models (e.g., based on Lidar or radar) for detecting driving situations are also feasible. Light or radio detection and ranging methods proved valuable for Situation Detection in difficult visual conditions (Vargas et al., 2021). Yet, some AV manufacturers like Tesla[4] employ vision-based Situation Detection using only cameras. They argue that automated driving is a computer vision problem, inspired by how human drivers rely only on their eyesight. In this sense, they argue that humans also do not need detailed point clouds (often in centimeters) to safely drive in scenarios with varying visual complexity (e.g., during fog, in a crowded street, or at a large junction) and, therefore, more sensors (e.g., for Lidar or radar) do not necessarily imply improved AV driving performance. Instead, more sensors might generate more entropy in the system, as sensor data has to be considered by the driving automation, increasing the complexity. Likewise, we argue that using only vision-based models to visualize AVs' Situation Detection and Prediction is realistic regarding current and future AV technology. However, as cameras are prone to visual distractions (e.g., fog or interfering light sources) (Vargas et al., 2021; Dietmayer, 2016), vision-based models might provide limited detection and prediction in some driving situations. This limitation is also visible in the converted user-chosen videos (see open feedback in Section 6.0.1).

The detail of Lidar/radar point clouds capable of replicating the environment in 3D could compensate for the limitations of vision-based models. Although Lidar/radar may not be needed for AV Trajectory Planning, users could have more trust in a system with such an extensive understanding of scenarios. Therefore, we assume that additional visualizations like Lidar point clouds could increase trust and user acceptance. In this case, the Lidar and radar systems would be only used as visualizations for the user and are not part of the functional implementation of AVs.

Considering the benefits and drawbacks of these methods from the functionality perspective, future research should investigate the visualizations of vision-based and Lidar/radar-based approaches regarding their effects on user acceptance and trust in AVs. Both methods have value in different application areas and should, therefore, not be neglected due to policies or industry decisions by some manufacturers.

7.5. Visualizing Uncertainty and Effects on Trust

We focused especially on the visualization of uncertainty by showing the current representation of the internal state of the neural networks. This led, as some participants mentioned, to flickering over time. Generally, there are two approaches to conveying uncertainty: graphical annotations and visual encodings. There are five types of uncertainty encodings for figures: fuzziness, location, arrangement, size, and transparency (Padilla, Kay and Hullman, 2022).

[4] `https://www.tesla.com/support/transitioning-tesla-vision`; Accessed: 29.02.2024

These can also be combined. MacEachren, Roth, O'Brien, Li, Swingley and Gahegan (2012) additionally includes color (hue, saturation, and value), shape, orientation, and grain. They report that fuzziness, location, value, and arrangement work well to visualize uncertainty. Size and transparency were deemed potentially usable compared to saturation, which was ranked low (MacEachren et al., 2012). As values and their associated uncertainty value are often mapped independently (Correll, Moritz and Heer, 2018; Kay, 2019), Value-Suppressing Uncertainty Palettes have been proposed (Correll et al., 2018). This allows more values to be communicated with low uncertainty and fewer values with high uncertainty. In the proposed visualization approach, we encoded the uncertainty in a temporal fashion by directly showing the changing world representation of the AV (i.e., via the flickering). Therefore, we argue that this specific uncertainty visualization applies to the AV context. Including a different visualization coding, although interesting (e.g., blurring, transparency, or value-suppressing uncertainty palettes), would have introduced even more flickering, which was already perceived as distracting.

Van Der Bles, van der Linden, Freeman and Spiegelhalter (2020) evaluated the effect of communicating uncertainty in information. For example, they either only provided a numerical value (e.g., "between April and June 2017, the number of unemployed people in the United Kingdom was an estimated 1,484,000." (Van Der Bles et al., 2020, p. 2)) or they added numerical or textual information regarding the information limits (e.g., numerical "minimum 1,413,000 to maximum 1,555,000" (Van Der Bles et al., 2020, p. 2); or verbal: "The report states that there is some uncertainty around this estimate, it could be somewhat higher or lower" (Van Der Bles et al., 2020, p. 2)). They showed that communicating limits of information leads to greater uncertainty by the information receiver. However, this only led to a small decrease in trust. While we also saw a small, non-significant decrease in trust for experiment B with the visualization (see Table 2), the opposite occurred in experiment A (small, non-significant increase; see Table 1). We interpret this as such that the visualization of AV functionality supports the user in understanding the AV but that the trust was already appropriate to the functionality. While it would be possible that the visualization was not sufficient to alter trust, previous work already indicated that these visualizations are an adequate communication medium (Colley et al., 2020, 2021a, 2022).

7.6. EduLicit and the Need to Educate the Public

According to the European Charter for Researchers[5], the researcher's duty is to make their research and the results available to the public. However, there seems to be a lack of dissemination in various research areas, including health sector (Kerner, Rimer and Emmons, 2005) but also HCI (Smith, Nevarez and Zhu, 2020). This dissemination task is also challenging, as Smith et al. (2020) showed. In line with Smith et al. (2020), we argue that researchers should be able to communicate with the public directly. Our openly available website and the EduLicit method can be a way to advance this dissemination.

In our specific case of visualizing functionalities of AVs, previous work showed different effects of feedback visualizations in AVs. Colley et al. stated that it is "unlikely that a manufacturer will provide these visualizations to potential customers as these would also show potential inadequacies" (Colley et al., 2022, p. 17). Also, Liu et al. (2022) stated that "dispelling public misconceptions about AVs is necessary" (Liu et al., 2022, p. 1). Therefore, we built a website that shows a subset of these (i.e., Situation Detection and pedestrian intention as a subset of Situation Prediction). This is in accordance with work by Dixon (2020) who state that "capabilities [functionalities] of automation are often overstated" (Dixon, 2020, p. 1). Accordingly, our results show that the participants were highly surprised by the results of the visualizations. Moreover, our results, notably from the open feedback and the qualitative analysis of uploaded videos (subsubsection 6.0.1 and Section 6.0.2), directly highlight the public's misconceptions and knowledge gaps regarding AV functionalities and limitations. For instance, feedback such as 'Interesting that the car sees pedestrians in a window of a shop...' and concerns over misidentifications (e.g., construction posts detected as pedestrians) illustrate the public's engagement and learning process through our platform (the Educate part). Such insights emphasize the effectiveness of hands-on, interactive tools in conveying the complexities of AV technology to a non-expert audience. Moreover, tools that enable users to (re-)experience the recorded scenarios in 3D driving environment replications (e.g., see (Jansen, Britten, Häusele, Segschneider, Colley and Rukzio, 2023)) would facilitate further insights. The diverse scenarios uploaded by participants, alongside their feedback (the Elicit part), not only enriched our understanding of public perceptions but also validated EduLicit as a method for AV education.

By directly confronting and clarifying the 'false' detections and limitations of current AVs (or state-of-the-art neural networks), we provide a realistic perspective on the technology's readiness and areas for improvement. While integrating multiple cameras and other sensors with additional temporal coherence (in contrast, we currently evaluate

[5] https://euraxess.ec.europa.eu/jobs/charter/european-charter; Accessed: 29.02.2024

the video frame by frame) could lead to more visually pleasing and more accurate predictions, we argue that our approach still reflects state-of-the-art AV functionalities. Based on the participant responses, we also see the need to provide tools to the public to understand the limitations of AVs (Colley et al., 2022), which our website implementation of EduLicit is a first step towards that direction.

7.7. Limitations and Future Work

The demographic information of experiment A shows that approximately 26% of our participants were self-employed, which is more than double the nationwide average in the USA in August 2023[6]. Also, the share of students (approx. 15%) is higher than the national average in 2020[7]. The required participant number was determined a priori and is sufficient based on comparisons to the determined effect sizes in previous work by Colley et al. (2021a) and (Colley et al., 2021b) (see the comparison between Kendall effect size and Cohen's d[8]. However, we only focused on subjective dependent measures. Therefore, an interpretation of the results should consider higher variance.

In experiment B, participants were mostly younger males. Therefore, the transferability of the results to other age groups is unclear. Also, there was a large share of college students (50%) who may be more tech-savvy than the general population and, thus, might have a deeper understanding of AVs. The sample size was relatively small (N=22). However, recruitment attempts via online survey portals (e.g., prolific.co) had yet to be successful due to the unreliability of participants, probably caused by the high video recording efforts. Experiment B used monocular vision-based models to visualize AVs' Situation Detection and Prediction. Although this was also done by Colley et al. (2022) and true to the functional implementation levels of AVs (Dietmayer, 2016), we cannot infer that our visualizations based on vision-based models correspond to visualizations of real-world algorithms. Besides, the used models were not optimized to run on user-chosen videos with different viewpoints, light reflection from the vehicle's windshield, shaking user hands while recording, and varying camera focus, which may have resulted in erroneous detection and prediction results. Finally, in Experiment B, participants rated their own uploaded videos. This could be a source of bias in their judgment.

On our website, we leveraged the open-sourced approaches by Wu et al. (2019) and Mordan et al. (2021) for Situation Detection and Situation Prediction. Machine learning is a fast-moving field, and novel approaches appear daily. Therefore, our approach should be re-used with newer approaches. For Situation Detection, this could be Segment Anything (Ke, Ye, Danelljan, Tai, Tang, Yu et al., 2024) or InternImage (Wang, Dai, Chen, Huang, Li, Zhu, Hu, Lu, Lu, Li et al., 2023).

When publicly available, future studies should also apply a monocular ego-AV trajectory planning visualization on user-uploaded videos in line with the simulation in experiment A. In addition, the overall fidelity of the visualizations could be improved by a stricter pre-selection of user-uploaded videos aimed at reducing erroneous detection and predictions. Future studies should also increase the number of participants and consider countries with different driving laws and cultures to enhance understanding of users' perceptions towards imperfect Situation Detection and Prediction.

In our study, participants could upload one video and be informed about the AV functionalities once. However, this might be insufficient if people discover new potential challenges for AVs over a long time frame (e.g., weeks and months) that could alter their perceptions of this technology. Thus, future work should investigate the longitudinal effects of using the EduLicit website for educating about AV functionalities.

8. Conclusion

Potential users' understanding of AV technology significantly lacks crucial information, necessitating efforts to improve awareness and knowledge. In this work, we employed EduLicit, a novel method to elicit scenarios that users deem challenging while serving as a platform for informing users about how well these AVs perceive their environment (Situation Detection and Situation Prediction) to educate the general public. In the first experiment (N=76), we successfully replicated the results of a previous study (Colley et al., 2022) regarding the effects of uncertainty visualization in AVs. We conducted a follow-up experiment (N=22) using a custom website that employs the EduLicit method. Participants could upload videos of self-chosen challenging scenarios to our website, which then automatically visualized the results of the two neural networks for Situation Detection and Prediction. Most user-chosen scenarios included multiple vulnerable road users and/or large intersections. The users' scenarios, subjective

[6] `https://www.bls.gov/news.release/pdf/empsit.pdf`; Accessed: 29.02.2024
[7] Back-to-school statistics; Accessed: 29.02.2024
[8] `https://cran.r-project.org/web/packages/effectsize/vignettes/interpret.html`; Accessed: 11.04.2024

ratings, and open feedback emphasized their skepticism toward detecting vulnerable road users (e.g., pedestrians) and predicting their intentions. Our website enabled us to identify scenarios that the public perceives as challenging for AVs and, at the same time, to educate the public about their functionalities. Our study highlights the importance of public education in the field of AV technology. It suggests that future research should include components that help demystify the technology for the general public.

Open Science

Upon acceptance, we will open-source our website and backend code, thereby enabling other researchers to leverage our EduLicit approach.

ACKNOWLEDGMENTS

We thank all study participants.

CRediT authorship contribution statement

Pascal Jansen: Conceptualization, Methodology, Software, Hardware, Investigation, Data Curation, Writing- Original Draft, Writing- Reviewing and Editing, Visualization, Project administration. **Mark Colley:** Conceptualization, Methodology, Software, Hardware, Investigation, Formal Analysis, Data Curation, Writing- Original Draft, Writing- Reviewing and Editing, Visualization. **Tim Pfeifer:** Software, Methodology, Investigation. **Enrico Rukzio:** Resources, Supervision, Writing- Reviewing and Editing.

References

Administration, N.H.T.S., 2023. Traffic safety facts 2020: A compilation of motor vehicle crash data (Report No. DOT HS 813 375). https://crashstats.nhtsa.dot.gov/Api/Public/ViewPublication/813375. [Online; accessed 07-NOVEMBER-2023].

Avetisyan, L., Ayoub, J., Zhou, F., 2022. Investigating explanations in conditional and highly automated driving: The effects of situation awareness and modality. Transportation Research Part F: Traffic Psychology and Behaviour 89, 456–466. URL: https://www.sciencedirect.com/science/article/pii/S1369847822001589, doi:https://doi.org/10.1016/j.trf.2022.07.010.

Beller, J., Heesen, M., Vollrath, M., 2013. Improving the driver–automation interaction: An approach using automation uncertainty. Human Factors 55, 1130–1141. URL: https://doi.org/10.1177/0018720813482327, doi:10.1177/0018720813482327, arXiv:https://doi.org/10.1177/0018720813482327. pMID: 24745204.

Brooke, J., et al., 1996. Sus-a quick and dirty usability scale. Usability evaluation in industry 189, 4–7.

Cai, Y., Jing, P., Wang, B., Jiang, C., Wang, Y., 2023. How does "over-hype" lead to public misconceptions about autonomous vehicles? a new insight applying causal inference. Transportation research part A: policy and practice 175, 103757.

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, Association for Computing Machinery, New York, NY, USA. p. 181–191. URL: https://doi.org/10.1145/3409120.3410648, doi:10.1145/3409120.3410648.

Colley, M., Eder, B., Rixen, J.O., Rukzio, E., 2021a. 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, Association for Computing Machinery, New York, NY, USA. URL: https://doi.org/10.1145/3411764.3445351.

Colley, M., Krauss, S., Lanzer, M., Rukzio, E., 2021b. How should automated vehicles communicate critical situations? a comparative analysis of visualization concepts. Proc. ACM Interact. Mob. Wearable Ubiquitous Technol. 5. URL: https://doi.org/10.1145/3478111, doi:10.1145/3478111.

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. Proc. ACM Interact. Mob. Wearable Ubiquitous Technol. 6. URL: https://doi.org/10.1145/3534609, doi:10.1145/3534609.

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. Proc. ACM Interact. Mob. Wearable Ubiquitous Technol. 7. URL: https://doi.org/10.1145/3631408, doi:10.1145/3631408.

Correll, M., Moritz, D., Heer, J., 2018. Value-suppressing uncertainty palettes, in: Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems, Association for Computing Machinery, New York, NY, USA. p. 1–11. URL: https://doi.org/10.1145/3173574.3174216, doi:10.1145/3173574.3174216.

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, Association for Computing Machinery, New York, NY, USA. URL: https://doi.org/10.1145/3411764.3445575.

Czaja, S.J., Charness, N., Fisk, A.D., Hertzog, C., Nair, S.N., Rogers, W.A., Sharit, J., 2006. Factors predicting the use of technology: findings from the center for research and education on aging and technology enhancement (create). Psychology and aging 21, 333.

Dietmayer, K., 2016. Predicting of machine perception for automated driving, in: Autonomous Driving. Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 407–424.

Dixon, L., 2020. Autonowashing: The greenwashing of vehicle automation. Transportation Research Interdisciplinary Perspectives 5, 100113. URL: https://www.sciencedirect.com/science/article/pii/S2590198220300245, doi:https://doi.org/10.1016/j.trip.2020.100113.

Du, M., Zhang, T., Liu, J., Xu, Z., Liu, P., 2022. Rumors in the air? exploring public misconceptions about automated vehicles. Transportation research part A: policy and practice 156, 237–252.

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, Association for Computing Machinery, New York, NY, USA. p. 1–14. URL: https://doi.org/10.1145/3313831.3376484, doi:10.1145/3313831.3376484.

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.

Filiz, C., 2020. Can autonomous vehicles prevent traffic accidents?, in: Accident Analysis and Prevention. IntechOpen, Rijeka, Croati.

Flohr, L.A., Valiyaveettil, J.S., Krüger, A., Wallach, D.P., 2023. Prototyping autonomous vehicle windshields with ar and real-time object detection visualization: An on-road wizard-of-oz study, in: Proceedings of the 2023 ACM Designing Interactive Systems Conference, Association for Computing Machinery, New York, NY, USA. p. 2123–2137. URL: https://doi.org/10.1145/3563657.3596051, doi:10.1145/3563657.3596051.

Galloway, A., Caudwell, C., 2018. Speculative design as research method: From answers to questions and "staying with the trouble", in: Undesign. Routledge, pp. 85–96.

Ghazizadeh, M., Lee, J.D., Boyle, L.N., 2012. Extending the technology acceptance model to assess automation. Cognition, Technology & Work 14, 39–49.

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. URL: https://www.sciencedirect.com/science/article/pii/S1369847820304587, doi:https://doi.org/10.1016/j.trf.2020.06.021.

Hales, D., 2013. Design fictions an introduction and provisional taxonomy. Digital Creativity 24, 1–10.

Hart, S.G., Staveland, L.E., 1988. Development of nasa-tlx (task load index): Results of empirical and theoretical research, in: Advances in psychology. Elsevier, Amsterdam, The Netherlands. volume 52, 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, Association for Computing Machinery, New York, NY, USA. p. 210–217. URL: https://doi.org/10.1145/2516540.2516554, doi:10.1145/2516540.2516554.

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.

Holländer, K., Colley, M., Rukzio, E., Butz, A., 2021. A taxonomy of vulnerable road users for hci based on a systematic literature review, in: Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems, Association for Computing Machinery, New York, NY, USA. URL: https://doi.org/10.1145/3411764.3445480.

Holthausen, B.E., Wintersberger, P., Walker, B.N., Riener, A., 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, Association for Computing Machinery, New York, NY, USA. p. 40–47. URL: https://doi.org/10.1145/3409120.3410637, doi:10.1145/3409120.3410637.

Innes, J.M., Fraser, C., 1971. Experimenter bias and other possible biases in psychological research. European Journal of Social Psychology 1, 297–310.

Jansen, P., Britten, J., Häusele, A., Segschneider, T., Colley, M., Rukzio, E., 2023. Autovis: Enabling mixed-immersive analysis of automotive user interface interaction studies, in: Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems, pp. 1–23.

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. Proc. ACM Interact. Mob. Wearable Ubiquitous Technol. 6. URL: https://doi.org/10.1145/3534617, doi:10.1145/3534617.

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.

Kay, M., 2019. How much value should an uncertainty palette suppress if an uncertainty palette should suppress value? statistical and perceptual perspectives.

Ke, L., Ye, M., Danelljan, M., Tai, Y.W., Tang, C.K., Yu, F., et al., 2024. Segment anything in high quality. Advances in Neural Information Processing Systems 36.

Kerner, J., Rimer, B., Emmons, K., 2005. Introduction to the special section on dissemination: dissemination research and research dissemination: how can we close the gap? Health Psychology 24, 443.

Kim, G., Yeo, D., Jo, T., Rus, D., Kim, S., 2023. What and when to explain? on-road evaluation of explanations in highly automated vehicles. Proc. ACM Interact. Mob. Wearable Ubiquitous Technol. 7. URL: https://doi.org/10.1145/3610886, doi:10.1145/3610886.

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 (IJIDeM) 9, 269–275.

Körber, M., 2019. Theoretical considerations and development of a questionnaire to measure trust in automation, in: Bagnara, S., Tartaglia, R., Albolino, S., Alexander, T., Fujita, Y. (Eds.), Proceedings of the 20th Congress of the International Ergonomics Association (IEA 2018), Springer International Publishing, Cham. pp. 13–30.

Kunz, F., Nuss, D., Wiest, J., Deusch, H., Reuter, S., Gritschneder, F., Scheel, A., Stübler, M., Bach, M., Hatzelmann, P., Wild, C., Dietmayer, K., 2015. Autonomous driving at ulm university: A modular, robust, and sensor-independent fusion approach, in: 2015 IEEE Intelligent Vehicles Symposium (IV), IEEE. IEEE, New York, NY, USA. pp. 666–673. doi:10.1109/IVS.2015.7225761.

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, Association for Computing Machinery, New York, NY, USA. p. 164–175. URL: https://doi.org/10.1145/3239060.3239074, doi:10.1145/3239060.3239074.

Kunze, A., Summerskill, S.J., Marshall, R., Filtness, A.J., 2019. Conveying uncertainties using peripheral awareness displays in the context of automated driving, in: Proceedings of the 11th International Conference on Automotive User Interfaces and Interactive Vehicular Applications, Association for Computing Machinery, New York, NY, USA. p. 329–341. URL: https://doi.org/10.1145/3342197.3344537, doi:10.1145/3342197.3344537.

Kyriakidis, M., Happee, R., de Winter, J.C., 2015. Public opinion on automated driving: Results of an international questionnaire among 5000 respondents. Transportation research part F: traffic psychology and behaviour 32, 127–140.

Lanzer, M., Stoll, T., Colley, M., Baumann, M., 2021. Intelligent mobility in the city: the influence of system and context factors on drivers' takeover willingness and trust in automated vehicles. Frontiers in Human Dynamics 3, 42. doi:10.3389/fhumd.2021.676667.

Lee, J.D., See, K.A., 2004. Trust in automation: Designing for appropriate reliance. Human factors 46, 50–80.

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.

Löcken, A., Frison, A.K., Fahn, V., Kreppold, D., Götz, M., Riener, A., 2020. Increasing user experience and trust in automated vehicles via an ambient light display, in: 22nd International Conference on Human-Computer Interaction with Mobile Devices and Services, Association for Computing Machinery, New York, NY, USA. URL: https://doi.org/10.1145/3379503.3403567, doi:10.1145/3379503.3403567.

Löcken, A., Heuten, W., Boll, S., 2016. Autoambicar: Using ambient light to inform drivers about intentions of their automated cars, in: Adjunct Proceedings of the 8th International Conference on Automotive User Interfaces and Interactive Vehicular Applications, Association for Computing Machinery, New York, NY, USA. p. 57–62. URL: https://doi.org/10.1145/3004323.3004329, doi:10.1145/3004323.3004329.

MacEachren, A.M., Roth, R.E., O'Brien, J., Li, B., Swingley, D., Gahegan, M., 2012. Visual semiotics & uncertainty visualization: An empirical study. IEEE transactions on visualization and computer graphics 18, 2496–2505.

Makransky, G., Petersen, G.B., 2021. The cognitive affective model of immersive learning (camil): A theoretical research-based model of learning in immersive virtual reality. Educational Psychology Review , 1–22.

Mayer, R.C., Davis, J.H., Schoorman, F.D., 1995. An integrative model of organizational trust. Academy of management review 20, 709–734.

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. doi:10.1109/TITS.2021.3107587.

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.

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), IEEE, New York, NY, USA. pp. 194–199. doi:10.1109/ARSO51874.2021.9542835.

Padilla, L., Kay, M., Hullman, J., 2022. Uncertainty visualization. Wiley, Hoboken, NJ, USA. chapter 22. pp. 405–421.

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, Association for Computing Machinery, New York, NY, USA. p. 340–348. URL: https://doi.org/10.1145/3543174.3546976, doi:10.1145/3543174.3546976.

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, Association for Computing Machinery, New York, NY, USA. p. 91–99. URL: https://doi.org/10.1145/3012709.3012735, doi:10.1145/3012709.3012735.

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.

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, Association for Computing Machinery, New York, NY, USA. URL: https://doi.org/10.1145/3411764.3446647.

Smith, C.E., Nevarez, E., Zhu, H., 2020. Disseminating research news in hci: Perceived hazards, how-to's, and opportunities for innovation, in: Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems, Association for Computing Machinery, New York, NY, USA. p. 1–13. URL: https://doi.org/10.1145/3313831.3376744, doi:10.1145/3313831.3376744.

Taş, Ö.Ş., Kuhnt, F., Zöllner, J.M., Stiller, C., 2016. Functional system architectures towards fully automated driving, in: 2016 IEEE Intelligent vehicles symposium (IV), IEEE. IEEE, New York, NY, USA. pp. 304–309.

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

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

Van Der Bles, A.M., van der Linden, S., Freeman, A.L., Spiegelhalter, D.J., 2020. The effects of communicating uncertainty on public trust in facts and numbers. Proceedings of the National Academy of Sciences 117, 7672–7683.

Vargas, J., Alsweiss, S., Toker, O., Razdan, R., Santos, J., 2021. An overview of autonomous vehicles sensors and their vulnerability to weather conditions. Sensors 21, 5397.

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, IEEE, New York, NY, USA. pp. 14408–14419.

Wiegand, G., Eiband, M., Haubelt, M., Hussmann, H., 2020. “i’d like an explanation for that!”exploring reactions to unexpected autonomous driving, in: 22nd International Conference on Human-Computer Interaction with Mobile Devices and Services, Association for Computing Machinery, New York, NY, USA. URL: https://doi.org/10.1145/3379503.3403554, doi:10.1145/3379503.3403554.

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, Association for Computing Machinery, New York, NY, USA. p. 105–107. URL: https://doi.org/10.1145/3379336.3381502, doi:10.1145/3379336.3381502.

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. URL: https://doi.org/10.1162/pres_a_00320, doi:10.1162/pres_a_00320, arXiv:https://doi.org/10.1162/pres_a_00320.

Wintersberger, P., Nicklas, H., Martlbauer, T., Hammer, S., Riener, A., 2020. Explainable automation: Personalized and adaptive uis to foster trust and understanding of driving automation systems, in: 12th International Conference on Automotive User Interfaces and Interactive Vehicular Applications, Association for Computing Machinery, New York, NY, USA. p. 252–261. URL: https://doi.org/10.1145/3409120.3410659, doi:10.1145/3409120.3410659.

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, Association for Computing Machinery, New York, NY, USA. p. 119–130. URL: https://doi.org/10.1145/3543174.3545253, doi:10.1145/3543174.3545253.

Wu, Y., Kirillov, A., Massa, F., Lo, W.Y., Girshick, R., 2019. Detectron2. https://github.com/facebookresearch/detectron2.

A. Ratings Analyst A

Table 3
Analyst A. Passenger Cars=Cars. Nonjunction=NJ. Non-Occupants=N-Occ. Light Truck=LT. Large Truck=LaT. Pedalcyclists=Cyc.. Pedestrians=Ped.

Video	Weather	Lighting	Types	Types	Types	Junction	Junction	# Lanes	# Lanes	N-Occ	N-Occ	Speed Limit	Viewpoint	Location	Remarks
C11N09M00	Rain	Daylight	Car	LT		NJ		One		Ped	Cyc.	50	Ped	Ulm Hbf	short after rain
E11N09M00	Normal	Daylight	Bus			NJ		Two		Ped		50	Ped	Uni Ulm	
E20B08F99	Normal	Daylight	Car	LT	LaT	Junction-Intersection	NJ	Two				70	Passenger	Biberach	
J11K11M97	Normal	Daylight	Car	LaT		NJ		Two		Ped	Cyc.	100	Passenger	prob. Biber-ach	with construction
L25W04M99	Normal	Daylight	Car	Bus	LT	Junction-Intersection		Two		Ped	Cyc.	50	Ped	Biberach	
M12K04F00	Normal	Daylight	Car	Bus	LT	NJ		Two		Ped		50	Ped	Biberach Hbf	
M23B11M93	Normal	Daylight	Car	LT	LaT	Junction-Intersection	NJ	Two		Ped		50	Passenger	Oberdischingen	
P25K02F01	Normal	Daylight	Car	LT		NJ		Two				50	Ped		
S23M12M99	Normal	Daylight	Car	LT	LaT	NJ		One		Ped		50	Ped	Ulm Hbf	
T26H03M98	Normal	Daylight	Car	LT		Junction-Intersection		Two				50	Passenger		
A11D09F00	Normal	Daylight	Car	LT		NJ		Four	Two	Ped	Cyc.	50	Passenger		
B10K05F00	Normal	Daylight	Car	Motorcycle		NJ		Two		Ped		50	Passenger		
D18S07M00	Normal	Daylight	Car	LT		NJ		Two		Ped	Cyc.	50	Passenger		
H15P05M63	Normal	Daylight	Car	LaT	LT	NJ		Two				100	Passenger		overtaking
K08P05M02	Normal	Daylight	Car	LT	Bus	Junction-Intersection		Four		Ped	Cyc.	50	Dashcam	Great Britain	Dashcam - crazy
L18P09M99	Normal	Dark (lighted)	Car	LT		NJ		Three				120	Passenger		
M27K10F40	Normal	Daylight	Car	LT	LaT	NJ		Two		Ped	Cyc.	50	Passenger		
M28H04M87	Normal	Daylight	Car	LT	LaT	Junction-Intersection		Four		Ped	Cyc.	50	Ped		
S06P11F66	Normal	Daylight	Car			Junction-Intersection		Four		Ped		60	Passenger		
S28S07M03	Normal	Daylight	Car	LT	LaT	Junction-Intersection		Two	Four	Ped			Dashcam		Multiple videos online
V13M02F97	Normal	Daylight	Car	LT	LaT	NJ		Two				120	Passenger		
S23M10M95	Normal	Daylight	Car	LT	LaT	Junction-Intersection	NJ	Three	Three			100	Passenger		

B. Ratings Analyst B

Table 4
Video codings from Analyst B. Passenger Cars=Cars.

Video	Weather	Lighting	Types	Types	Types	Junction	Junction	# Lanes	# Lanes	N-Occ	N-Occ	Speed Limit	Viewpoint	Location	Remarks
C11N09M00	Rain	Daylight	Car	LT		NJ		One		Ped	Cyc.	50	Ped	Ulm Hbf	
E11N09M00	Normal	Daylight	Bus			NJ		Two		Ped		50	Ped	Uni Ulm	
E20B08F99	Normal	Daylight	Car	LT	LaT	Junction-Intersection-Related	NJ	Two				70	Passenger	Biberach, near Jordanbad	
J11K11M97	Normal	Daylight	Car	LaT		NJ		Two		Ped	Cyc.	100	Passenger	Near Ulm	
L25W04M99	Normal	Daylight	Car	Bus	LT	Junction-Intersection		Two		Ped	Cyc.	50	Ped	Biberach, near Jordanbad	
M12K04F00	Normal	Daylight	Car	Bus	LT	NJ		Two		Ped		50	Ped	Biberach Hbf	
M23B11M93	Normal	Daylight	Car	LT	LaT	Junction-Intersection	NJ	Two		Ped		50	Passenger	Oberdischingen	
P25K02F01	Normal	Daylight	Car	LT		NJ		Two				50	Ped		
S23M12M99	Normal	Daylight	Car	LT	LaT	NJ		One		Ped		50	Ped	Ulm Hbf	
T26H03M98	Normal	Daylight	Car	LT		Junction-Intersection		Two				50	Passenger		
A11D09F00	Normal	Daylight	Car	LT		NJ		Four	Two	Ped	Cyc.	50	Passenger		
B10K05F00	Normal	Daylight	Car	Motorcycle		NJ		Two		Ped		50	Passenger		
D18S07M00	Normal	Daylight	Car	LT		NJ		Two		Ped	Cyc.	50	Passenger		
H15P05M63	Normal	Daylight	Car	LaT	LT	NJ		Two				100	Passenger		
K08P05M02	Normal	Daylight	Car	LT	Bus	Junction-Intersection		Four		Ped	Cyc.	50	Dashcam	Great Britain	
L18P09M99	Normal	Dark (lighted)	Car	LT		NJ		Three				120	Passenger		
M27K10F40	Normal	Daylight	Car	LT	LaT	NJ		Two		Ped	Cyc.	50	Passenger		
M28H04M87	Normal	Daylight	Car	LT	LaT	Junction-Intersection		Four		Ped	Cyc.	50	Ped		
S06P11F66	Normal	Daylight	Car			Junction-Intersection		Four		Ped		60	Passenger		
S28S07M03	Normal	Daylight	Car	LT	LaT	Junction-Intersection		Two	Four	Ped			Dashcam		
V13M02F97	Normal	Daylight	Car	LT	LaT	NJ		Two				120	Passenger		
S23M10M95	Normal	Daylight	Car	LT	LaT	Junction-Intersection	NJ	Three	Three			100	Passenger		