Robot-like In-vehicle Agent for a Level 3 Automated Vehicle with Emotions

XINGJIAN ZENG, Eindhoven University of Technology, The Netherlands
MD SHADAB ALAM*, Eindhoven University of Technology, The Netherlands
PAVLO BAZILINSKYY, Eindhoven University of Technology, The Netherlands

In-vehicle agents (IVAs) have emerged as a transformative innovation for intelligent transportation systems. This paper presents the development and evaluation of a robot-like IVA prototype with emotional feedback capabilities for SAE Level 3 automated vehicles. A user study assessed emotional interactions between drivers and IVA. The results showed that emotional feedback and driver working status did not have a significant effect on average workload or acceptance (usefulness and satisfaction). However, emotional feedback influenced physical and temporal demands, and its interaction with working status significantly affected the overall workload. Voice communication remained the main interaction mode, especially when drivers were engaged in other tasks. The study highlighted the challenges of accurately detecting emotions through facial recognition in automated driving scenarios, emphasised the need to consider physical conditions such as fatigue and stress, and insight into the participants' perspectives towards the IVA robot.

Additional Key Words and Phrases: In-Vehicle Agent, Robot-Like Agent, Emotional feedback, Emotion detection, Emotive driving

ACM Reference Format:
Xingjian Zeng, Md Shadab Alam, and Pavlo Bazilinskyy. 2025. Robot-like In-vehicle Agent for a Level 3 Automated Vehicle with Emotions. In *Proceedings of Make sure to enter the correct conference title from your rights confirmation email (Conference acronym 'XX).* ACM, New York, NY, USA, 15 pages. https://doi.org/XXXXXXX.XXXXXXX

1 Introduction

With advancements in automotive technology (AT) and artificial intelligence (AI), in-vehicle agents (IVAs) have become pivotal innovations within intelligent transportation systems (ITS). IVAs, typically serving as driving assistants integrated within vehicle operational systems, are broadly classified into voice agents, virtual agents, and physical agents. The primary objective is to assist the driver with driving tasks, thereby enhancing the overall driving experience [21]. Among these types, physical IVAs have gained increasing attention due to their intuitive interactive capabilities, the potential for stronger emotional connection, and deep integration with smart cockpit technologies (https://www.nio.com/nl_NL/el8).

The first physical IVA, the Affective Intelligent Driving Agent (AIDA), was introduced by the MIT Media Lab in 2013 [35]. AIDA used facial expressions displayed on a screen and provided navigation assistance through interactive communication. Similarly, Carvatar [37] further employed facial expressions to convey information and thus enhance the driver's trust. More recently, the Robot Human-Machine Interface (RHMI) [30] incorporated dynamic visual cues, such as eye colours and body movements, to provide timely takeover request warnings. Furthermore, Srivatsan et al.

*Corresponding Author

Authors' Contact Information: Xingjian Zeng, esse.zeng009@gmail.com, Eindhoven University of Technology, Eindhoven, The Netherlands; Md Shadab Alam, m.s.alam@tue.nl, Eindhoven University of Technology, Eindhoven, The Netherlands; Pavlo Bazilinskyy, p.bazilinskyy@tue.nl, Eindhoven University of Technology, Eindhoven, The Netherlands.

Fig. 1. Benchmark of Existing and Proposed IVAs.

[2] demonstrated the potential of robotic agent gestures to significantly improve passenger experience in automated vehicles (AVs).

The commercial adoption of physical IVAs has notably advanced, particularly in Asian countries with successful examples such as NIO's NOMI (https://youtube.com/watch?v=SAZ2Dd9lrVc), Baidu's Xiaodu (https://developer.apollo.auto/platform/dueros_cn.html), and Japan's Mochi (https://dasai.com.au/pages/meet-mochi). These agents typically feature geometric designs combined with digital screens that display facial expressions. However, their current implementations largely rely on voice commands or virtual interaction methods without incorporating gestures, which somewhat limits their potential for richer interactions. Figure 1 shows some of the existing and proposed IVAs.

Recent trends in automated vehicles have emphasised the driving experience as a key element of vehicle design. Innovations such as smart cockpits with personalised screens (https://aito.auto/), integrated entertainment system purposes (https://www.nio.com/nl_NL) and integrated entertainment systems indicate a shift towards more immersive user experiences https://aito.auto/model/m9/, (https://www.nio.com/nl_NL/el8). Concurrently, research indicates that future IVAs will increasingly adopt implicit interactions, leveraging embodied AI and multimodal interfaces to engage directly with users rather than simply facilitating communication between users and the vehicle [7, 29]. For example, IVAs could automatically adjust the driving style based on detected passenger emotions, thereby enhancing comfort and satisfaction [29].

Currently, interactions with IVAs remain predominantly explicit, involving voice commands or physical buttons. This approach limits their potential for deeper engagement and user experience improvements. Recognising this limitation, the present research explores the possibilities of implicit interaction between drivers and physical IVAs through facial emotion detection. Specifically, this project proposes the integration of emotional recognition capabilities into physical IVAs to foster richer, context-sensitive interactions, significantly improving the overall driving experience.

1.1 Related work

1.1.1 Benefits of In-Vehicle Agents. In the manual driving context, the IVAs can not only help with driving-related tasks like vehicle-to-vehicle communication (both vehicles need to install IVA) [8], or non-driving related tasks like comfort children to reduce distractions for the driver [6], but also minimise driver's distraction by decreasing the number of directed utterances with a set of robots [14], reduce driver's fatigue through social communication [18], and mitigate drivers' negative affective status through giving positive comments about the situation [23].

IVAs can explain the system status and intentions of an automated vehicle (AV) [10, 16, 22, 26, 37]. The user interface (UI) of IVAs can be a voice UI [16], a visual UI [10], or a physical UI [2]. Lee and Jeon [21] suggest that physical agents aid in better driving behaviour and overall experience, especially in the context of automated driving (AD). Zihsler et al. [37] and Chakravarthi et al. [2] showed that physical agents with facial expressions and gestures, respectively, can increase trust in AVs.

In an AD context, IVAs perform better in improving overall experience [21], such as explaining the status of the system with animation of a chauffeur avatar and a world in miniature [10], or using the "How + Why message" to lead better driving performance [16]. On the other hand, an IVA can serve as a companion by adopting a conversational dialogue style, using emotional tones and first-person language, which fosters a 'human-agent relationship' with the driver [19], giving the driver a sense of a 'human-agent relationship'. Furthermore, IVAs can increase trust and acceptance in AD using social cues and anthropomorphism to translate the state of the vehicle into human behaviour and expressions, which can be intuitively interpreted by the driver [37].

1.1.2 Interaction with In-Vehicle Agents. Voice interaction is a common communication method for IVAs in SAE Level 3 AD vehicles due to its minimal visual distraction [33]. Research on IVA voice interaction, including speech emotion and gender, indicates that no single voice suits all listeners and situations [12]. Lee et al. found that voice agents aligning with social role stereotypes (informative male and social female) enhance perceived ease of use (PEU) and perceived usefulness (PU) [20]. Jeon et al. showed the effectiveness of an in-vehicle software agent in mitigating effects on driver situation awareness and performance [11]. Ruijten et al. showed that conversational interfaces are more trusted, liked, anthropomorphised, and perceived as more intelligent than graphical UIs [26].

As IVAs evolve from voice-only agents to physical agents, interactions become more complex. Both virtual and physical agents can engage in visual interactions, with virtual agents being 2D or 3D characters, and physical agents having a physical appearance and facial expressions [7, 10, 15, 30, 34, 37]. However, the interesting thing is that except for AIDA published in 2014, other concepts are all in the context of AD.

Gestures are a unique feature of physical agents compared to other agents. The robot developed by Srivatsan et al. [2] shows that robotic objects are a promising technology to improve passengers' experience in AVs. RHMI developed by Tanabe et al. [30] can adjust the turning angle, speed, and opening angle of the lid to inform different levels of emergency: normal state, unstable state, and suspended state.

Social interactions, such as small talk, significantly increase driver trust compared to voice interactions alone [17]. Although robot agents can be visually distracting, yet increase trust, voice agents are preferred in low-speed situations [33]. Drivers have mixed attitudes towards conversational robot agents [22]. Both voice and robot agents improve likability and perceived warmth, with voice agents better at anthropomorphism, and robot agents offering greater competence and lower workload [32].

1.1.3 Implicit Interactions Through Facial Emotion Detection. The concept of implicit interactions was introduced by Ju & Leifer [13], emphasising that interactions should not rely solely on explicit user commands, but should adapt based on context and environment. This approach reduces cognitive load and enhances the user experience. Recent studies suggest that implicit interaction could be applied to highly automated vehicles [29] and would probably be the key feature of future IVAs [7]. Stampf et al. (2022) proposed five key input modalities used to recognise driver or passenger states and intentions: physiological signals, auditory signals, visual signals (gaze, blink, and pupil dilation), kinaesthetic characteristics (including facial) and profile [29]. However, if we despise the profile input (driving behaviour and preference could need long-term data input) and extract facial features, the other five input modalities overlap with the five ways Garcia-Garcia et al. [5] indicated to detect emotion: emotion from speech, emotion from text, emotion from facial expressions, emotion from body gestures and movements, and emotion from physiological states. This means that, in the case of AD scenarios, the input modalities could be regarded as different aspects of emotion detection.

However, only facial expressions and speech are suitable for driving scenarios above these choices because the driver and passengers' body gestures and movements are greatly restricted during driving, and the V-touch (https:

//youtube.com/watch?v=opkaJcPS7s8) proposed by Kia is still a concept. However, if the driver is driving alone without talking, it is impossible to detect the driver's emotion from speech. However, facial expressions contain more messages (55%) than voice intonation (38%) and spoken words (7%) [24], and can be detected constantly as long as the driver's face is captured. In addition, Ekman & Dacher (1997) argued that facial expressions are universal and provide sufficient information to predict emotions in his exploratory work [4], which is another advantage of facial expressions over speech (different languages may need to translate).

The general computational flow for facial emotion recognition contains five steps: image or data acquisition step, mage preparation (preprocessing) step, feature selection/extraction step, classification step, results and validation procedure [1]. Today, the facial recognition project could have extremely high accuracy in facial recognition tasks, such as deepface (https://github.com/serengil/deepface) (97.53%) and the highest accuracy of 98. 83% using the stationary wavelet transform for facial emotion recognition, which allows emotion detection in driving scenarios.

1.2 Aim of Study

This study is based on previous work [36]. In that project, it was found that both the interactions of facial expressions and gestures can reduce the workload and increase usefulness and satisfaction. However, gestures appear to be more functional and preferred by the driver, while facial expressions seem to be more emotional and preferred by passengers. The present study aims to explore the implicit interactions between the driver and the IVA robot through the detection of facial emotions and the corresponding feedback. Therefore, two research questions are defined: **RQ1:** *What impact would a robot-like in-vehicle agent equipped with emotional feedback have on a Level 3 driving experience when the driver is working/available?* and **RQ2:** *Which ways of interaction (voice, facial expression, gesture) in emotional feedback is first noticed when the driver is working/available?* In the context of this work, AD is assumed to be SAE Level 3 [27]. In this project, a robot-like IVA with emotional feedback was designed and developed to answer these two questions.

2 Design Process

2.1 Concept

The final concept sketch, which was created after two iterations, is shown in Figure 2. The soft LED (model: HRP-2064) was chosen to display the IVA's facial expressions. A gear transmission was used instead of the rubber wheel transmission because the gear transmission was more stable and easy to assemble. Due to safety concerns, the final prototype can be easily locked to prevent secondary damage in car accidents. On the other hand, the behaviour designed for the IVA robot was updated into two parts: the first part is to help with driving tasks and explain the current state (Table 1), and the other part is emotional feedback (Table 2). Table 1 was based on the table of IVA behaviour in seven highway scenarios [36]; they have the same structure, while Table 1 adds a shutdown scenario to complete the overall user experience and new gestures. Table 2 is based on the driver's emotions detected by the camera (Raspberry Pi Camera Module 2) and analysed through DeepFace (https://github.com/serengil/deepface), (OpenFace (https://github.com/TadasBaltrusaitis/OpenFace) was replaced because the video stream formats of the camera and OpenFace cannot match each other). There are seven emotions in the DeepFace library, including angry, fear, neutral, sad, disgust, happy, and surprise, which is quite similar to the 7 basic emotions (https://www.paulekman.com/universal-emotions/) (except for neutral replacing contempt). For each emotion detected, the IVA robot has the corresponding feedback (including dialogue, facial expressions, and gestures).

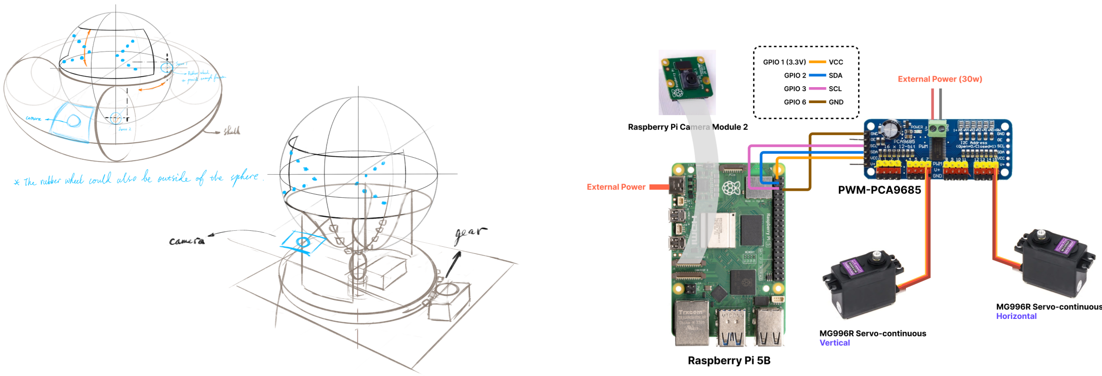


Fig. 2. Sketch of the final concept (left) and circuit (right).

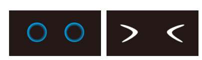
Scenarios	**Dialogue**	**Facial expressions (5)**	**Gestures H & V**
Wake up	—"Welcome! My name is Eva. Shall we start our trip?" –Yes — "Here we go!"	 Neutral Anticipation	Up 50°(half speed)
Enter highway	"We will enter highway ahead."	 Happy	Left 60°, wait 1s, right 60°(half speed)
Speed limit and speed report	"The speed limit is 90km/h, and right now we are at 87km/h."	 Happy	Left 60°, wait 1s, right 60°(half speed)
Overtaking	—"The front car is driving too slow, we gonna overtake it! wow Nice!"	 Neutral Anticipation	Left 80°, right 80°, left 80°, right 80°(full speed)
Lane changing (construction)	"Seems there is a construction ahead, we need to change lane."	 Surprise	Up 50°, wait 0.5s, left 60°, wait 1s, right 60°, wait 0.5s, down 50°(half speed)
Congestion	"Seems there is traffic jam ahead, we need to slow down."	 Sadness	Down 50°, wait 0.5s, left 60°, right 60°, wait 0.5s, up 50°(half speed)
Exit highway	"We will exit the highway ahead."	 Happy	Left 60°, wait 1s, right 60°(half speed)
Shut down	"Thanks for your participation, and wish you a lovely day!"	 Happy	down 50°(half speed)

Table 1. Updated IVA behaviour (gestures, facial expressions, and dialogues) in eight highway scenarios.

Emotion detected (7)	Dialogue	Facial expressions	Gestures H & V
Happy	"It's great to see you in such a good mood! Would you like to share what's making you happy?"	 Happy	Left 60°, right 60°(half speed)
Neutral	(None)	 Neutral	(None)
Surprise	"Everything is under control. I'm here if you need support."	 Surprise	Up 50°, left 60°, right 60°, down 50°(half speed)
Sad	"It seems like you're feeling a bit down. Would talking or some relaxing music help?"	 Neutral	Left 50°, right 50°(half speed)
Angry	"Traffic can be tough. Don't worry, the system is optimized to handle it. How about some calming music?"	 Happy	Left 60°, right 60°(half speed)
Fear	"It seems like you're feeling a bit uneasy. I'll monitor everything and react to keep us safe. Is there anything I can do for you?"	 Happy	Left 60°, right 60°(half speed)
Disgust	"Is there something unpleasant I can help with? Perhaps a window adjustment or a different temperature setting?"	 Neutral	Left 60°, right 60°(half speed)

Table 2. IVA behaviour of emotional feedback.

2.2 Materials

A Raspberry Pi 5 (model: B, 4GB RAM, 32GB SD, Raspberry Pi OS (64bit): Bookworm 12.8); a PWM-PCA9685(16-channel I2C PWM-Servo Controller - PCA9685); two servomotors (MG996R Servo - 10kg - Continuous); a camera (Raspberry Pi Camera Module 2); a soft LED (model: HRP-2064); an AC-DC external power supply for servomotors (model: AED45US05, 5V, 6.0A); 3D print (see supplementary material, should include "gear", "gearb", "head", "motorsupport", "shellbuttongear", "shellsupport", "sphering"); laser cut MDF board (see supplementary material, should contain one file for 9mm and one file for 2mm).

2.3 Hardware

The assembly began with the integration of the electronic components (Figure 2). Two servomotors were connected to the PWM-PCA9685 via Dupont wires, using Pin 7 for pitch and Pin 15 for yaw control. Communication between the Raspberry Pi and the PCA9685 was established through the following GPIO pins: GPIO 1 - VCC, GPIO 2 - SDA, GPIO 3 - SCL, and GPIO 6 - GND. The camera module was connected through the Raspberry Pi CAM 1 port. External power was

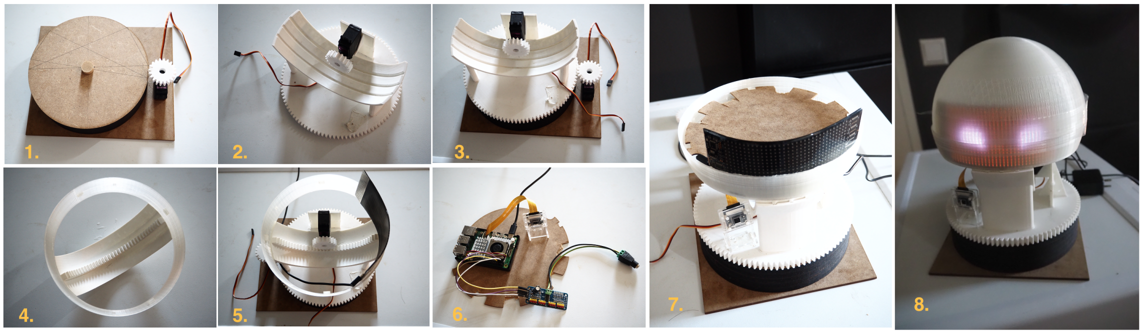


Fig. 3. Hardware assembly process (from 1 - 8).

supplied to the Raspberry Pi, the soft LED module, and the PWM-PCA9685 (Note: an external power source is required for the PCA9685 if the total current of the servomotors exceeds 500 mA).

After completion of the electronics setup, the mechanical structure was assembled (Figure 3). Two 2mm MDF boards were bonded together, and two small gears were fixed to the servomotors. Five 9mm thick and one 2mm thick large circular MDF boards were glued and placed on the marked area, with a 9mm MDF cylinder fixed at the centre to serve as the rotation axis for the "gearb." The MDF stand was then completed. The "motorsupport," camera, and "shellsupport" were mounted onto "gearb." The yaw servomotor (Pin 15) was attached to the MDF stand, while the pitch servomotor (Pin 7) was mounted on the "motorsupport." The "shellbuttongear" and "spherering" were assembled, with the soft LED installed along the internal curve of the "spherering." Finally, the Raspberry Pi, attached on a 2mm MDF platform, was placed inside the "spherering," and the "head" was secured on top.

2.4 Software

The software environment was based on the Raspberry Pi OS (64-bit, Bookworm 12.8). Source code is available in supplementary material. All emotional feedback is set in the config.py according to Table 2, including voice, facial expressions, and gestures. Other basic behaviour according to Table 1 are written in mainbasic.py. All audio was generated from PlayHT (https://play.ht) (see supplementary material, file names beginning with "d" are the audios for emotional feedback). However, sometimes the audio will just skip around 1 second when playing with Bluetooth connection, so all audio of basic behaviour is edited to add 1 second of silence at the beginning, using FFmpeg (https://www.ffmpeg.org).

The facial expressions of the IVA robot are actually uploaded through another way, since the soft LED cannot connect directly to the GPIO of Raspberry Pi, which means the soft LED cannot be controlled directly by Raspberry Pi. However, it has a mobile phone app and uses Bluetooth to send data, which makes it possible for a laptop to receive data sent from the mobile app to the soft LED. First, connect the mobile phone and laptop and the soft LED. Then send the expressions from the mobile app, and the data is a long string of numbers and letters. Next, connect the soft LED with the Raspberry Pi and use the Raspberry Pi to send the same string of numbers and letters; the expressions will eventually be shown on the soft LED. That is where the strings of numbers in config.py come from.

3 Experiment

In order to answer *RQ*1 and *RQ*2, an experiment was conducted to evaluate the final concept design. The experiment had two groups: Group *W* (working) and Group *N* (no working). Since SAE Level 3 AD (https://www.sae.org/standards/content/j3016_202104) is conditionally automated, working here means all actions that need to be immersed and highly focused for an uncertain duration, such as reading a book, chatting, or watching short videos on a mobile phone. In order to control the variables, participants in Group *W* need to play the Poly Bridge game (https://apps.apple.com/us/app/poly-bridge/id1197552569) on an iPad (model: A2229) during the tasks, while participants in Group *N* cannot use any personal products on digital screens (including smart phones). Each participant in Group *W* and Group *N* needs to complete two tasks: Task *B* (basic) and Task *E* (emotional feedback). The behaviour of the IVA robot in Task *B* is shown in Table 1 , and the code is in mainbasic.py. Task *E* is based on Task *B*, adding an emotional feedback function according to Table 2 , and the code is in main.py. The study was approved by the Ethics Review Board of Eindhoven University of Technology, and the participants gave their informed consent to use their data.

3.1 Setup and Preparation

Figure 4 shows the experimental setup. A JBL GO 3 (model: JBLG03BLK) was connected to the Raspberry Pi inside the prototype, and the Raspberry Pi was remotely controlled (https://github.com/raspberrypi/documentation/tree/develop/documentation/asciidoc/computers/remote-access) by the laptop (Apple Macbook A2442). An RCA 32" LED TV (model: RS32F3) was also connected to the laptop via an HDMI cable. The detailed parameters of the layout are shown in Figure 4. The participants were asked to sit in front of the centre of the RCA TV, and the prototype should face the participants with an angle of 45° so that the camera could detect the emotions of the participants. The position of the prototype is settled on the front right of the participant, corresponding to the position above the dashboard in a real car.

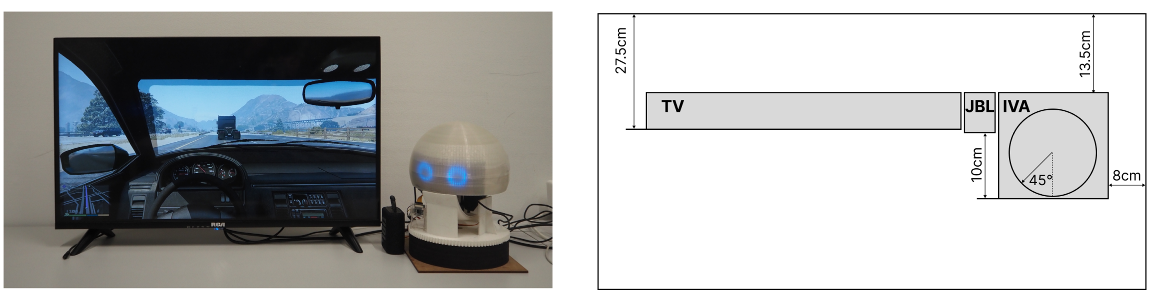


Fig. 4. Scenario setup (left) and layout of the equipment (right).

The scenarios videos were recorded in the GTA V video game (https://www.rockstargames.com/zh/gta-v) (running on a Windows PC according to Table 1, and the highway route is chosen from downtown to Beeker's Garage. To get an inside view of AD, two mods were applied: (1) Dynamic Vehicle First Person Camera Mod (https://youtube.com/watch?v=jwxgmAHtwIY), allowing the camera inside the vehicle to get the driver's perspective and (2) Enhanced Native Trainer Mod https://youtube.com/watch?v=UHHXTh0Xdow, which makes characters invisible (i.e., no hands holding the steering wheel were visible, providing a sense of driving in an AV). The videos were then edited together into one video (see supplementary material).

In Task *E*, the prototype will not only execute the behaviour in Table 1 but also the emotional feedback in Table 2. In case two actions interfere with each other and avoid frequent actions, emotional detection and feedback were only enabled after the following scenarios: welcome, speed report, construction, and traffic jam. Feedback can only be triggered once for each scenario. Each behaviour in Task B was matched to the video.

Furthermore, the servomotor (MG996R Servo—10kg—Continuous) can only use speed and time as parameters. Even when using mathematical methods to set the angle as the parameter in servomotor.py, errors still occurred due to the startup delay. Consequently, the prototype cannot return to the exact starting position after each task. Thus, the position of the prototype needs to be adjusted before each task according to Figure 4.

3.2 Participants

A total of 12 participants (age: M = 26.33, SD = 2.50; 7 females and 5 males) from Eindhoven University of Technology joined the user test through the user test link posted on the social media platform. And no financial incentives were offered for the user test. All participants were over 18 years of age and had a driver's licence. Three participants had experience driving an automated vehicle (with automatic lane changing / automatic turning, automatic overtaking / automatic navigation, or higher functions, such as Tesla, Waymo, Apollo, etc.). One has been a passenger of an automated vehicle.

3.3 Procedures

First, participants signed the consent form and completed an online survey (see supplementary material) to collect basic information. The slides then briefly introduced background information about the project to the participants (see supplementary material). The participants then took a seat and had two tasks to complete: Task *B* and Task *E*. To reduce the error, the sequence of Task *E* and Task *B* is changed for each half of the participants in Group *W* and Group *N*. The prototype was controlled by the Raspberry Pi, which was remote controlled by the author on the laptop during the experiment. After each task, participants were asked to fill in the NASA Task Load Index scale [9] on the official App (https://apps.apple.com/us/app/nasa-tlx/id1168110608) to measure workload and the acceptance scale [31] (integrated in the online survey) to measure the overall experience on the iPad, since Ju & Leifer suggested that implicit interaction could reduce cognitive load and enhance user experience. Finally, a semi-structured interview was conducted to collect the experiment experience. During each task, participants were asked to imagine themselves acting as drivers in a SAE Level 3 AD (which means the driver can use automated driving conditionally and still needs to be responsible for the driving duration) and to have an automated driving experience with the prototype. Participants in Group *W* will be asked to play the Poly Bridge game (https://apps.apple.com/us/app/poly-bridge/id1197552569) on an iPad (model: A2229) during the tasks, while participants in Group *N* cannot use any personal products on digital screens (including smart phones). All participants were allowed to look up and check the situation at any time. If they felt that they wanted to take over control immediately, they were asked to inform the author about it.

4 Method of Data Analysis

There are four types of raw data collected during the experiment from each participant: basic information survey (including Driving Behaviour Questionnaire), NASA TLX Scale on mobile App (https://apps.apple.com/us/app/nasa-tlx/id1168110608) [9], online acceptance scale [31], and a semi-structured interview (see supplementary material).

However, the results exported from the NASA TLX App contain too much useless information and the csv files of Rating Score and Pairwise after each task were exported separately. Thus, every participant produced 4 files (2 for Rating Score, 2 for Pairwise), and the raw data of NASA TLX contain 48 files (24 for Rating Score, 24 for Pairwise, see the folder named "nasa"). Since the data of Rating Score and Pairwise are stored in different csv files, rerfine.py was created to scan all 48 csv files, extract all the useful information, and merge them into two new .csv files (nasa_rs_data.csv for Rating Score, and nasa_pw_data.csv for Pairwise). Then, use Excel to remove the "," and fill the blank cells with the value 0. After that, using the code in nasatlx_analysis.py extract rows of data according to Group and Task and calculate the results in the table exported (Results_from_nasa_tlx.csv).

Since Google Form (https://www.google.com/forms/about/) can only set values larger than 0 on a linear scale (see supplementary material), all values in the cells need to subtract 2 and mirror the values according to the guideline [3] to get acceptance.csv. Then, use the code in acceptance_analysis.py to calculate the results and export them into a new table (Results_from_acceptance_scale.html).

Furthermore, SPSS can only accept a wide format of CSV to perform the repeated measures ANOVA test (https://www.ibm.com/docs/en/spss-statistics/30.0.0). Alpha value of 0.05 was used. In this case, results of NASA TLX [9] and acceptance scale [31] were adjusted from long format to wide format. The experiment interview was analysed through thematic analysis. The themes were generated after coding the data in the transcription.

5 Results

5.1 NASA TLX

Table 3 shows the mean and standard deviation values of each workload dimension score and the weighted average score of the overall workload of different groups (*W* or *N*) and different Tasks (*E* or *B*). Since the standard deviations shown in Table 3 were too large, a statistical analysis is needed in this case. Table 3 also shows the analysis that examined the effects of different working status (W/N) and emotional feedback (E/B) on average weighted workload measures using the NASA Task Load Index [9]. There were two independent variables in the experiment: emotional feedback (E/B) as a within-groups variable, and working status (W/N) as a between-groups variable, which means two-way ANOVA should be replaced by repeated measures ANOVA in this case. Thus, the repeated measures ANOVA test was applied with SPSS (see supplementary material). The main results are as follows: (1) The between-subjects factor working status (W/N) did not have a significant effect on the average weighted workload measured by NASA TLX ($p=0.139$); (2) The within-subjects factor emotional feedback (E/B) also did not have a significant effect on the average weighted workload measured by NASA TLX ($p=0.057$); (3) The interaction effect between emotional feedback and working status was significant for the average weighted workload measured by NASA TLX ($p=0.037$).

5.2 Acceptance Scale

The results of the acceptance scale are shown in Table 4. Similarly to NASA TLX, the standard deviations shown in Table 4 were too large, and statistical analysis is needed. Table 4 also shows that the analysis examined the effects of different working status (W/N) and emotional feedback (E/B) on overall usefulness and satisfaction measures using the Van der Laan Acceptance Scale [31]. There were two independent variables in the experiment: emotional feedback (E/B) as a within-groups variable, and working status (W/N) as a between-groups variable, which means two-way ANOVA should be replaced by repeated measures ANOVA in this case. Thus, the repeated measures ANOVA test was applied with SPSS (see supplementary material). The results suggest that emotional feedback (E/B) and working status did not

Table 3. Results from the NASA TLX Scale [9].

	W Task *E* M(SD)	*W* Task *B* M(SD)	*N* Task *E* M(SD)	*N* Task *B* M(SD)			
Mental Demand (%)	49 (32)	55 (31)	52 (20)	33 (20)			
Physical Demand (%)	16 (27)	21 (25)	22 (27)	16 (19)			
Temporal Demand (%)	66 (30)	48 (30)	43 (29)	25 (19)			
Performance (%)	45 (27)	68 (24)	14 (12)	13 (13)			
Effort (%)	50 (31)	41 (17)	22 (17)	20 (13)			
Frustration (%)	33 (14)	39 (20)	21 (15)	16 (15)			
					Repeated-measures ANOVA		
					W/N	E/B	W/N*E/B
Average (%)	47 (26)	47 (23)	31 (22)	22 (17)	p=0.139	p=0.057	p=0.037

Note: *W*=working group, *N*=no-working group.

Table 4. Results from the Acceptance Scale [31].

Negative (-2)	Positive (+2)	*W* Task *E* M(SD)	*W* Task *B* M(SD)	*N* Task *E* M(SD)	*N* Task *B* M(SD)			
Useless	Useful	0.67 (1.03)	1.00 (1.10)	1.17 (0.75)	1.33 (0.82)			
Unpleasant	Pleasant	0.83 (0.98)	0.67 (1.03)	1.00 (0.89)	1.33 (0.52)			
Bad	Good	0.83 (0.75)	1.00 (0.63)	1.17 (0.75)	1.33 (0.52)			
Annoying	Nice	0.83 (0.75)	0.83 (0.75)	1.17 (0.75)	1.33 (0.52)			
Superfluous	Effective	0.83 (0.75)	0.67 (0.82)	0.83 (0.75)	1.00 (1.26)			
Irritating	Likeable	0.67 (0.82)	0.50 (0.84)	1.17 (0.75)	1.33 (0.52)			
Worthless	Assisting	1.17 (0.41)	1.33 (0.82)	1.33 (0.52)	1.33 (0.52)			
Undesirable	Desirable	0.67 (1.03)	0.17 (1.17)	1.33 (0.52)	1.33 (0.52)			
Sleep-inducing	Raising Alertness	0.33 (1.21)	0.00 (1.55)	1.00 (0.89)	0.67 (0.52)			
						Repeated-measures ANOVA		
						W/N	E/B	W/N*E/B
Overall usefulness score		0.77 (0.64)	0.80 (0.63)	1.10 (0.53)	1.13 (0.47)	p=0.299	p=0.801	p=1.000
Overall satisfaction score		0.75 (0.79)	0.54 (0.80)	1.17 (0.61)	1.33 (0.44)	p=0.090	p=0.926	p=0.412

Note: *W*=working group, *N*=no-working group.

significantly impact the perception of Usefulness or Satisfaction. Furthermore, there was no significant interaction between emotional feedback and working status. These findings suggest that within the given experimental conditions, emotional feedback and working status did not play a significant role in shaping user acceptance levels.

5.3 Interview

The experiment interview was analysed through thematic analysis. The themes were generated after coding the data in the transcription. Details can be found in the supplementary material. The first column contains the participants' group and the number of participants who contributed to each code. There are five themes (technology concern, efficiency, attitude, prototype, and experiment) according to 16 codes extracted from the interview transcription (see supplementary material).

The results indicate that most of the participants trust the robot (P1, P3, P4, P5, P6, P7, P8, P9, P10, P12) while others have doubts about the AD technology (P2, P5, P11) and the accuracy of facial detection (P1, P2, P3, P4, P5, P7, P8, P10, P12). Furthermore, some suggested that other emotions (physical conditions) like tiredness and stress should also be taken into consideration (P6, P9, P11). Although the IVA robot behaviour is designed for basic highway situations, and emotional feedback can encourage communication between the driver and IVA robot (P2, P6, P7, P11), it can also make the participant annoyed (P2, P5, P7, P9) and distracted (P2, P5, P7). Even the IVA robot itself can be a good reminder of the situations (P1, P4, P7), some of the dialogue would be time-consuming (P2, P3) or cause misunderstanding (P4, P11). Therefore, some participants prefer direct suggestions (P3, P6) rather than open questions. Besides, some participants complained about the noise (P1, P2, P5, P9), and scenarios (P6, P8, P10). Other participants mind the meaning of the behaviour (P5, P7, P9) and the ideal size of the prototype (P1, P10).

6 Discussion

In this project, a robot-like IVA prototype with emotional feedback was developed for an SAE Level 3 AV and was evaluated in an experiment. Here are notable results: (1) Although the within-subjects factor emotional feedback (E/B) did not have a significant effect on the average weighted workload measured by NASA TLX ($P=0.057$), the statistical analysis (see supplementary material) indicated that the emotional feedback factor showed a significant influence on specific measures, particularly the dimension of physical demand ($F(1,10) = 5.052$, $p = 0.048$) and temporal demand ($F(1,10) = 5.213$, $p = 0.046$), but their impact varies depending on the nature of the task and the working status. Additionally, the interaction between emotional feedback and working status highlights the potential combined effects on task load, underscoring the importance of considering both factors in SAE Level 3 AD scenarios. However, due to the small sample size ($N=12$), the study results may be more susceptible to individual variability, reducing the stability of the data. (2) Within the given experimental conditions, emotional feedback and working status did not play a significant role in shaping user acceptance levels. Interestingly, Task *E* of both groups had higher scores in the dimension of Sleep-inducing - Raising Alertness than those of Task *B*, which means emotional feedback can make the driver more alert and sober, also supported by P9 in the interview. This may apply to some specific situations. For example, the driver is tired or stressed but the AD cannot be enabled. (3) The reactions and attitudes to emotional feedback are various due to different situations and individual preferences. Most felt annoyed (P2, P5, P7, P9) when working while others would like to chat with it when available (P2, P6, P7, P11). (4) The primary way to convey information to the driver is still through voice interaction; most participants could remember the dialogue clearly but could hardly tell the behaviour of the IVA robot. Some participants admitted this idea (P2, P4, P10). So the emotional connection could also be through voice intonation and spoken words [24]. (5) Tiredness and stress are rather physical conditions than emotions, so it would be hard to detect and need other methods to measure (since the accuracy of facial expression detection is quite low during the experiment). On the other hand, these are typical physical reactions, indicating that the driver is not suitable to drive, which should be taken into consideration. (6) Even though deepface (https://github.com/serengil/deepface) was a powerful framework for recognising facial expressions, it didn't execute the right output most of the time. Possible reasons are inappropriate parameters that cannot recognise complex facial expressions like yawns, or different colours of participants' clothes that interfere with the background colour (since sometimes no one is in front of the camera but it still produces fear or angry). Also, if facial recognition detects emotions, which are strongly influenced by light, it probably doesn't work when driving at night. A possible solution might be combining different ways of emotion detection [28] or input modalities [29].

6.1 Limitations and Future Work

The experiment used the RCA TV to display the video to simulate driving scenarios, which could have caused errors, as the participants could not feel the acceleration and deceleration of the vehicle. When asked if they trusted the IVA robot during the experiment, P2 and P8 mentioned that they were not in the real car and knew they were safe, so they trusted the IVA robot (see the supplementary material). Furthermore, the experiment process is Wizard of Oz, which means that the video would not change if the participants tried to take over the control, and the driving route is the same. In addition, the prototype is larger than the ideal size (head diameter around 15cm), which might cause errors in the evaluation. Due to the size of the soft LED and the two servomotors, if the size is smaller, the different parts will interfere with each other. However, for industrial production, all components could be customised so that the mechanical structure could be smaller but more meticulous, just like the Nomi (https://youtube.com/watch?v=SAZ2Dd9lrVc). The gear transmission could also be replaced by magnetic levitation and momentum wheels. Furthermore, the emotive driving concept is cool but abstract, and the feedback of the IVA (vehicle system) is based on what data it received rather than what happened in reality, which is the reason why all the dialogues of emotional feedback only provide general suggestions or comforting words.

Future studies could improve the reliability and external validity of the research results by increasing the sample size, testing in real AD scenarios, or conducting a larger controlled trial. Besides, there are only positive or neutral expressions for emotional feedback because when drivers' negative emotions are detected, it is not reasonable for the IVA robot to show a negative expression. As P9 mentioned, sometimes a bug can be dangerous. But maybe the IVA could also act like a weak robot [25] to ask for help in order to meet everyone's need to care for someone else, which might be interesting to research in the future. Furthermore, based on the feedback of being shocked by the big truck or contradictory driving habits (P6, P8) and the reaction of participants during the experiment, emotion detection can be combined with machine learning to form a customised IVA robot (AD system) according to their own driving habits. For example, if the driver had a surprised expression each time driving parallel to a truck, it would prevent them from getting close to the truck in the future. The same idea can be applied to the speed limit and the following distance. Finally, there will be a unique IVA robot for each customer, which is exactly the profile part of the input modalities used to recognise the states and intentions of the driver or passenger [29].

7 Conclusions

In summary, in this project, a robot-like in-vehicle agent with emotional feedback was developed to explore the implicit interaction between the driver and IVA, and the prototype was evaluated through an experiment to answer the two proposed research questions. For **RQ1**, results showed that neither emotional feedback nor working status has a significant impact on the average weighted workload in AD scenarios. Emotional feedback might have an impact on the dimension of Physical and Temporal Demand, and the interaction effect between emotional feedback and working status was significant for the average weighted workload. However, more experimental samples are still needed to confirm and refine the above conclusions. As for **RQ2**, voice is still the main interaction between the IVA robot and the driver, especially when the driver is working. The results of the experiment also revealed the challenges of using facial recognition to detect emotions in the driving scenario, other physical conditions like tiredness or stress that need to be taken into consideration, and participants' perspectives towards the IVA robot.

8 Supplementary Material

Surveys, STL files, analysis, materials used in the experiment, and raw data can be found at: https://www.dropbox.com/scl/fo/962bbgpgdjerz7mpa0ik4/AFk-31Schu0wxc1kYgtgUfE?rlkey=h6h9hee11r90uc03xozvly8pk&st=0ju24wjq.
The maintained source code for the hardware with detailed operational instructions is available at https://github.com/esse009/emotion-face.

References

[1] Felipe Zago Canal, Tobias Rossi Müller, Jhennifer Cristine Matias, Gustavo Gino Scotton, Antonio Reis de Sa Junior, Eliane Pozzebon, and Antonio Carlos Sobieranski. 2022. A survey on facial emotion recognition techniques: A state-of-the-art literature review. *Information Sciences* 582 (2022), 593–617.

[2] Srivatsan Chakravarthi Kumaran, Toam Bechor, and Hadas Erel. 2024. A Social Approach for Autonomous Vehicles: A Robotic Object to Enhance Passengers' Sense of Safety and Trust. In *Proceedings of the 2024 ACM/IEEE International Conference on Human-Robot Interaction*. 86–95.

[3] Dick de Waard. 2017. Acceptance Scale A simple scale to asses system acceptance. https://www.hfes-europe.org/accept/accept.htm

[4] Paul Ekman and Dacher Keltner. 1997. Universal facial expressions of emotion, Segerstrale U, P. *Nonverbal communication: Where nature meets culture* (1997), 27–46.

[5] Jose Maria Garcia-Garcia, Victor MR Penichet, and Maria D Lozano. 2017. Emotion detection: a technology review. In *Proceedings of the XVIII international conference on human computer interaction*. 1–8.

[6] Michal Gordon and Cynthia Breazeal. 2015. Designing a virtual assistant for in-car child entertainment. In *Proceedings of the 14th International Conference on Interaction Design and Children*. 359–362.

[7] Yijie Guo, Hanyang Hu, Zhuoran Zhai, Yao Lu, Xuezhu Wang, Yuanling Feng, Zhihao Yao, Zhenhan Huang, Yuan Yao, and Haipeng Mi. 2023. Designing Future In-Vehicle Assistants: Insights from User Imaginations and Experiences. In *Proceedings of the Eleventh International Symposium of Chinese CHI*. 508–520.

[8] Toshiyuki Hagiya and Kazunari Nawa. 2020. Acceptability evaluation of inter-driver interaction system via a driving agent using vehicle-to-vehicle communication. In *Proceedings of the 11th Augmented Human International Conference*. 1–8.

[9] Sandra G Hart and Lowell E Staveland. 1988. Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. In *Advances in psychology*. Vol. 52. Elsevier, 139–183.

[10] Renate Häuslschmid, Max von Buelow, Bastian Pfleging, and Andreas Butz. 2017. Supportingtrust in autonomous driving. In *Proceedings of the 22nd international conference on intelligent user interfaces*. 319–329.

[11] Myounghoon Jeon, Bruce N Walker, and Thomas M Gable. 2015. The effects of social interactions with in-vehicle agents on a driver's anger level, driving performance, situation awareness, and perceived workload. *Applied ergonomics* 50 (2015), 185–199.

[12] Marie Jonsson and Nils Dahlbäck. 2009. Impact of voice variation in speech-based in-vehicle systems on attitude and driving behaviour. *Human Factors and Ergonomics Society Europe Chapter (HFES)* (2009).

[13] Wendy Ju and Larry Leifer. 2008. The design of implicit interactions: Making interactive systems less obnoxious. *Design Issues* 24, 3 (2008), 72–84. doi:10.1162/desi.2008.24.3.72

[14] Nihan Karatas, Shintaro Tamura, Momoko Fushiki, and Michio Okada. 2018. Multi-party conversation of driving agents: the effects of overhearing information on lifelikeness and distraction. In *Proceedings of the 6th International Conference on Human-Agent Interaction*. 84–91.

[15] Nihan Karatas, Soshi Yoshikawa, Shintaro Tamura, Sho Otaki, Ryuji Funayama, and Michio Okada. 2017. Sociable driving agents to maintain driver's attention in autonomous driving. In *2017 26th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*. IEEE, 143–149.

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

[17] Johannes Maria Kraus, Florian Nothdurft, Philipp Hock, David Scholz, Wolfgang Minker, and Martin Baumann. 2016. Human after all: Effects of mere presence and social interaction of a humanoid robot as a co-driver in automated driving. In *Adjunct proceedings of the 8th international conference on automotive user interfaces and interactive vehicular applications*. 129–134.

[18] David R Large, Gary Burnett, Vicki Antrobus, and Lee Skrypchuk. 2018. Driven to discussion: engaging drivers in conversation with a digital assistant as a countermeasure to passive task-related fatigue. *IET Intelligent Transport Systems* 12, 6 (2018), 420–426.

[19] David R Large, Gary Burnett, and Leigh Clark. 2019. Lessons from Oz: design guidelines for automotive conversational user interfaces. In *Proceedings of the 11th International Conference on Automotive User Interfaces and Interactive Vehicular Applications: Adjunct Proceedings*. 335–340.

[20] Sanguk Lee, Rabindra Ratan, and Taiwoo Park. 2019. The voice makes the car: Enhancing autonomous vehicle perceptions and adoption intention through voice agent gender and style. *Multimodal Technologies and Interaction* 3, 1 (2019), 20.

[21] Seul Chan Lee and Myounghoon Jeon. 2022. A systematic review of functions and design features of in-vehicle agents. *International Journal of Human-Computer Studies* 165 (2022), 102864.

[22] Seul Chan Lee, Harsh Sanghavi, Sangjin Ko, and Myounghoon Jeon. 2019. Autonomous driving with an agent: Speech style and embodiment. In *Proceedings of the 11th international conference on automotive user interfaces and interactive vehicular applications: Adjunct proceedings*. 209–214.

[23] Shuling Li, Tingru Zhang, Wei Zhang, Na Liu, and Gaoyan Lyu. 2020. Effects of speech-based intervention with positive comments on reduction of driver's anger state and perceived workload, and improvement of driving performance. *Applied Ergonomics* 86 (2020), 103098.

[24] Albert Mehrabian. 2017. Communication without words. In *Communication theory*. Routledge, 193–200.

[25] Michio Okada. 2022. Weak robots. *JSAP Review* 2022 (2022), 220409. doi:10.11470/jsaprev.220409

[26] Peter AM Ruijten, Jacques MB Terken, and Sanjeev N Chandramouli. 2018. Enhancing trust in autonomous vehicles through intelligent user interfaces that mimic human behavior. *Multimodal Technologies and Interaction* 2, 4 (2018), 62.

[27] SAE. 2021. Taxonomy and definitions for terms related to driving automation systems for on-road motor vehicles. https://www.sae.org/standards/content/j3016_202104

[28] Mohammad Soleymani, Sadjad Asghari-Esfeden, Yun Fu, and Maja Pantic. 2015. Analysis of EEG signals and facial expressions for continuous emotion detection. *IEEE Transactions on Affective Computing* 7, 1 (2015), 17–28.

[29] Annika Stampf, Mark Colley, and Enrico Rukzio. 2022. Towards Implicit Interaction in Highly Automated Vehicles-A Systematic Literature Review. *Proceedings of the ACM on Human-Computer Interaction* 6, MHCI (2022), 1–21. doi:10.1145/3546726

[30] Hiroko Tanabe, Yuki Yoshihara, Nihan Karatas, Kazuhiro Fujikake, Takahiro Tanaka, Shuhei Takeuchi, Tsuneyuki Yamamoto, Makoto Harazawa, and Naoki Kamiya. 2022. Effects of a Robot Human-Machine Interface on Emergency Steering Control and Prefrontal Cortex Activation in Automatic Driving. In *International Conference on Human-Computer Interaction*. Springer, 108–123.

[31] Jinke D Van Der Laan, Adriaan Heino, and Dick De Waard. 1997. A simple procedure for the assessment of acceptance of advanced transport telematics. *Transportation Research Part C: Emerging Technologies* 5, 1 (1997), 1–10.

[32] Manhua Wang, Seul Chan Lee, Harsh Kamalesh Sanghavi, Megan Eskew, Bo Zhou, and Myounghoon Jeon. 2021. In-vehicle intelligent agents in fully autonomous driving: The effects of speech style and embodiment together and separately. In *13th International Conference on Automotive User Interfaces and Interactive Vehicular Applications*. 247–254.

[33] Manhua Wang, Seul Chan Lee, Genevieve Montavon, Jiakang Qin, and Myounghoon Jeon. 2022. Conversational voice agents are preferred and Lead to better driving performance in conditionally automated vehicles. In *Proceedings of the 14th International Conference on Automotive User Interfaces and Interactive Vehicular Applications*. 86–95.

[34] Kenton Williams, José Acevedo Flores, and Joshua Peters. 2014. Affective robot influence on driver adherence to safety, cognitive load reduction and sociability. In *Proceedings of the 6th international conference on automotive user interfaces and interactive vehicular applications*. 1–8.

[35] Kenton J Williams, Joshua C Peters, and Cynthia L Breazeal. 2013. Towards leveraging the driver's mobile device for an intelligent, sociable in-car robotic assistant. In *2013 IEEE intelligent vehicles symposium (IV)*. IEEE, 369–376.

[36] Xingjian Zeng, Md Shadab Alam, and Pavlo Bazilinskky. 2025. Robot-like in-vehicle agent for a level 3 automated vehicle. (2025). Preprint: https://shaadalam9.github.io/publications/xingjian_1.

[37] Jens Zihsler, Philipp Hock, Marcel Walch, Kirill Dzuba, Denis Schwager, Patrick Szauer, and Enrico Rukzio. 2016. Carvatar: increasing trust in highly-automated driving through social cues. In *Adjunct proceedings of the 8th international conference on automotive user interfaces and interactive vehicular applications*. 9–14.