

Queensland University of Technology
Brisbane Australia

This may be the author's version of a work that was submitted/accepted for publication in the following source:

Cockram, Lewis, Yu, Yueteng, Pardo Gaytan, Jorge, Li, Xiaomeng, Rakotonirainy, Andry, Kuo, Jonny, Demmel, Sebastien, Lenné, Michael G., & Schroeter, Ronald
(2026)
Small Talk, Big Impact? LLM-based Conversational Agents to Mitigate Passive Fatigue in Conditional Automated Driving.
In Oliver, Nuria, Shamma, David A., Candello, Heloisa, Cesar, Pablo, Lopes, Pedro, Bozzon, Alessandro, et al. (Eds.) *CHI '26: Proceedings of the 2026 CHI Conference on Human Factors in Computing Systems.*
Association for Computing Machinery (ACM), United States of America.

This file was downloaded from: https://eprints.qut.edu.au/263229/

© 2026 Copyright held by the owner/author(s).

This work is covered by copyright. Unless the document is being made available under a Creative Commons Licence, you must assume that re-use is limited to personal use and that permission from the copyright owner must be obtained for all other uses. If the document is available under a Creative Commons License (or other specified license) then refer to the Licence for details of permitted re-use. It is a condition of access that users recognise and abide by the legal requirements associated with these rights. If you believe that this work infringes copyright please provide details by email to qut.copyright@qut.edu.au

License: Creative Commons: Attribution 4.0

Notice: *Please note that this document may not be the Version of Record (i.e. published version) of the work. Author manuscript versions (as Submitted for peer review or as Accepted for publication after peer review) can be identified by an absence of publisher branding and/or typeset appearance. If there is any doubt, please refer to the published source.*

https://doi.org/10.1145/3772318.3790972

Small Talk, Big Impact? LLM-based Conversational Agents to Mitigate Passive Fatigue in Conditional Automated Driving

Lewis Cockram*
Queensland University of Technology
Brisbane, Australia
cockraml@qut.edu.au

Yueteng Yu*
Centre for Future Mobility
Queensland University of Technology
Brisbane, Australia
yueteng.yu@hdr.qut.edu.au

Jorge Pardo
Centre for Future Mobility
Queensland University of Technology
Brisbane, Australia
jorge.pardogaytan@qut.edu.au

Xiaomeng Li
Centre for Future Mobility
Queensland University of Technology
Brisbane, Australia
xiaomeng.li@qut.edu.au

Andry Rakotonirainy
Centre for Future Mobility
Queensland University of Technology
Brisbane, Australia
r.andry@qut.edu.au

Jonny Kuo
Seeing Machines
Melbourne, Australia
jonny.kuo@seeingmachines.com

Sebastien Demmel
Centre for Future Mobility
Queensland University of Technology
Brisbane, Australia
sebastien.demmel@qut.edu.au

Mike Lenné
Seeing Machines
Melbourne, Australia
mike.lenne@seeingmachines.com

Ronald Schroeter
Centre for Future Mobility
Queensland University of Technology
Brisbane, Australia
r.schroeter@qut.edu.au



Figure 1: Using a conversational agent to reduce passive fatigue in a L3 automated driving scenario.

*Both authors contributed equally to this research

This work is licensed under a Creative Commons Attribution 4.0 International License.

CHI '26, Barcelona, Spain

Abstract

Passive fatigue during conditional automated driving can compromise driver readiness and safety. This paper presents findings from

© 2026 Copyright held by the owner/author(s).

ACM ISBN 979-8-4007-2278-3/26/04
https://doi.org/10.1145/3772318.3790972

a test-track study with 40 participants in a real-world automated driving scenario. In this scenario, a Large Language Model (LLM) based conversational agent (CA) was designed to check in with drivers and re-engage them with their surroundings. Drawing on in-car video recordings, sleepiness ratings and interviews, we analysed how drivers interacted with the agent and how these interactions shaped alertness. Results show the CA is helpful for supporting vigilance during passive fatigue. Thematic analysis of acceptability further revealed three user preference profiles that implicate future intention to use CAs. Positioning empirically observed profiles within existing CA archetype frameworks highlights the need for adaptive design sensitive to diverse user groups. This work underscores the potential of CAs as proactive Human–Machine Interface (HMI) interventions, demonstrating how natural language can support context-aware interaction during automated driving.

CCS Concepts

• **Human-centered computing → User studies**.

Keywords

Human-Machine Interaction, Conversational AI, Automated Driving, Passive fatigue, Human-agent Collaboration, Attention management, Natural Language Interaction

ACM Reference Format:

Lewis Cockram, Yueteng Yu, Jorge Pardo, Xiaomeng Li, Andry Rakotonirainy, Jonny Kuo, Sebastien Demmel, Mike Lenné, and Ronald Schroeter. 2026. Small Talk, Big Impact? LLM-based Conversational Agents to Mitigate Passive Fatigue in Conditional Automated Driving. In *Proceedings of the 2026 CHI Conference on Human Factors in Computing Systems (CHI '26), April 13–17, 2026, Barcelona, Spain.* ACM, New York, NY, USA, 18 pages. https://doi.org/10.1145/3772318.3790972

1 Introduction

While automated vehicles (AVs) can reduce human error-related crashes, a paradox is introduced. By relieving drivers of manual control and reducing human error, SAE Level 3 (L3) [38] *Conditional Driving Automation* creates conditions for *passive fatigue* [47]. This state of cognitive underload, monotony, and vigilance decline undermines drivers' readiness to take over [24]. Addressing passive fatigue is not simply a technical challenge but a safety-critical issue in Human-AV interactions [11, 47].

Various countermeasures for passive fatigue have been proposed, such as non-driving-related tasks (NDRTs) [17, 22, 30, 48] or gamification [41] to stimulate driver engagement during automated driving. These can temporarily mitigate cognitive underload in automated driving [25]. Additionally, few can adapt to the context or needs of individual drivers [10]. This challenge can be conceptualised using *Malleable Attentional Resource Theory* (MART), which posits that attentional resources expand and contract in response to task demand [49]. Under low-demand conditions typical of automation, attentional resources contract, degrading vigilance and situational awareness, whereas overly demanding tasks can increase workload and induce active fatigue [24, 28]. Addressing the risks of underload situations is therefore as critical as mitigating those arising from excessive workload during control transitions. As a result, alternative forms of *implicit* intervention for AVs have gained attention. Leveraging natural response or social stimulation, implicit interventions can deliver subtle 'nudges' that guide users to maintain awareness unobtrusively [42].

Conversational agents (CAs) offer a novel and dynamic approach to delivering these nudges. CAs may engage users through natural social interaction with real-time, context-aware information about the environment[1]. Their design can draw on established archetypes, service-oriented, entertainment-focused, or socially companionable, each offering distinct ways to sustain engagement [21].

Despite CAs' potential, previous studies have largely relied on driving simulators or Wizard-of-Oz (WoZ) setups. This is due to limited access to AV-capable vehicles and the technical challenges of deploying real-time AI interaction [13, 18]. However, in intervention studies, simulators may suppress behavioural responses due to the absence of real motion and perceived risk, while WoZ setups constrain experimental control, making it difficult to replicate the precise timing and behavioural nuances of automated systems. Thus, the potential of real-time, large language model (LLM)-based agents in L3 automated driving remains unexplored [5]. This leaves unanswered questions about their effectiveness, usability, and acceptance in safety-critical contexts; a gap this paper aims to address.

Our study explores the following research questions (RQs):

- RQ1. How do users perceive the usage of a CA for supporting attention and engagement during monotonous automated driving?
- RQ2. What design and interaction factors shape users' perceptions of using a CA in the context of monotonous automated driving?
- RQ3. What effects do safety-oriented CA interventions have on drivers' behaviour and the states of alertness during authentic automated driving?

In this study, we introduced a real-time CA designed to maintain cognitive engagement during *automated driving*. 40 participants interacted with it during a monotonous, driving task on a closed test track. The voice-based conversations were short, contextually grounded, and aimed at supporting driver alertness without introducing distraction. Participants' alertness was assessed using the Karolinska Sleepiness Scale (KSS), alongside post-drive interviews to evaluate perceived usefulness, naturalness, and engagement.

This study contributes:

(1) One of the first field studies using a real CA, deployed in real-time, during *authentic* automated driving, demonstrating its capacity to mitigate passive fatigue and support driver alertness under highly ecologically valid conditions.
(2) Thematic user insights into drivers' experiences and perceptions, highlighting the factors that shape the acceptability of CAs in safety-critical contexts.
(3) Emerging user archetypes that align with established CA design frameworks, revealing how safety-first, entertainment-seeking, and socially oriented drivers value different aspects of conversational support.

Together, these contributions advance our understanding of how conversational interfaces can be designed to balance safety and engagement in highly automated vehicles. More broadly, they point to

[1] https://openai.com/index/hello-gpt-4o/

the value of archetype-sensitive design in shaping future human–AI mobility systems.

2 Related Work

2.1 Designing against Passive Fatigue in Conditional Automated Driving

In conditional automated driving, drivers' alertness may degrade during monotonous driving tasks. Prior work identifies two key dimensions of monotony, road design monotony and roadside variability, which can rapidly reduce driver alertness [19]. Such degradation widens the readiness gap between passive monitoring and active control [20]. This gap has motivated growing interest in Human–Machine Interface (HMI) strategies that can subtly reorient driver attention in preparation for potential transitions [2].

A variety of HMI solutions have been proposed, including explicit visual, auditory, and haptic alerts [15, 31, 45]. However, these modalities can disrupt the user experience, particularly when delivered during pre-transition periods where readiness is low but a takeover is not yet required. In response, researchers have increasingly explored "gentle" interaction methods in authentic driving contexts [51], including gamification and CAs.

Research has examined how playful or game-like elements can help keep drivers alert by adding entertainment value. Steinberger et al. [43] demonstrated how the inclusion of gamified elements can sustain driver attention during simulated driving. Introducing speed-control feedback, challenges and a score-based system was found to alleviate driver boredom and encourage engagement with the task. Bier et al. [3] also tested a gamified in-vehicle interaction system, asking trivia questions linked to the driving simulator task. They compared three groups: a control, a co-driver condition, and the gamified system. Results showed that both the co-driver and the gamified systems helped delay the onset of fatigue. The gamified system also led to better engagement over time.

These results point to the potential of leveraging lightweight and joyful interaction to maintain attention and engagement. Game-like tasks can increase cognitive load just enough to prevent boredom without overwhelming the driver. However, they also present design challenges. Systems need to be simple enough not to distract drivers from takeover events, but engaging enough to be meaningful.

Mahajan et al. [23] used a WoZ design to test whether a simple voice agent could reduce passive fatigue during a 30-minute L3 drive. The agent provided hazard alerts, mock reminders, and general conversation using pre-recorded messages every three minutes. Compared to driving without the agent, participants showed reduced sleepiness and no microsleeps. A follow-up study reported that agent-assisted participants were 39% more likely to complete a successful takeover [23]. This suggests that conversational interactions may both support alertness and improve takeover performance. Large et al. [17] reported similar effects in their WoZ study.

More recently, researchers such as Huang et al. [13] have begun introducing real-time LLM-powered CAs into a simulated driving task. Their system prompted drivers to engage in frequent, simple conversations during a long, monotonous session. High- and low-frequency conversations elicited greater electroencephalography (EEG) arousal than the absence of interaction. This study was one of the first to apply a genuine large LLM in a simulated driving context.

Most prior studies addressing passive fatigue, regardless of their intervention, used driving simulators, which do not fully reflect what happens when drivers are in motion [13, 17] on a road and are responsible for their own safety. Simulators are helpful for early testing, but do not capture all the cues that shape how people feel about in-vehicle technology. Studies have also shown that the presence or absence of vehicle motion itself can significantly affect driver responses [37]. Therefore, while past work shows that CAs and gamification can help interrupt passive fatigue, the effectiveness of playful, LLM-powered CAs in addressing passive fatigue *in a real-world driving context* remains unknown.

Our study addresses this gap with one of the first deployments of a fully-functional LLM-powered CA in a genuine L3 automated vehicle prototype during monotonous driving. The CA was designed to provide a natural, seamless and playful form of safety-forward interaction. This approach enabled us to gain rich insights into how participants experienced the agent strategy, how useful they perceived it to be, and what changes they would expect for future deployment.

3 Designing Zoe: A Conversational Agent to Combat Passive Fatigue

We designed a CA, Zoe, specifically to help drivers of conditionally automated vehicles combat passive fatigue. At the time of the study, Zoe was powered by the latest available model (OpenAI GPT-4[2] with a real-time API), ensuring responsiveness and naturalness in interaction. The authors acknowledge that other LLMs, as well as more recent versions of GPT, have since become available and may produce responses that differ from those generated during our study.

The design drew on prior simulator-based studies of conversational agents, which emphasised the importance of low-complexity dialogue, natural speech, and short communication in reducing fatigue [13, 17]. From a safety perspective, supporting situational awareness has been widely discussed in the context of L3 driving [7, 9], with particular emphasis on context-aware information exchange for developing adaptive interfaces and providing implicit interaction [4, 40].

Building on these findings, Zoe was developed around three core elements:

(1) **Engagement strategies** encouraged drivers to notice and describe their surroundings (e.g., "What can you see around us?") or comment on expected roadside wildlife.
(2) **Interaction management** protocols ensured the agent never assumed responses, paused if participants appeared busy, and included recovery strategies such as, "I apologise, could you please repeat that?"
(3) **Safety protocols** required the agent to regularly remind drivers that road attention always takes precedence and to immediately halt interaction if the CA became a distraction.

[2] https://openai.com/index/gpt-4-research/

In addition, light cognitive-weight gamification features were included to broaden appeal and encourage engagement. *Zoe* was prompted to provide context-relevant prompts tied to road vigilance. This included optional micro-interactions drawn selectively from the Entertainment archetype (e.g., brief challenges, environment-linked facts). The intention was to support attention and provide gamification elements [3] without impacting on safety goals. A little-shot prompting approach was used to enhance the user experience by reducing the risk of hallucination, improving the accuracy of personalised responses, and supporting greater stability, consistency, and clarity [29].

Design goals

(1) Safety first: safety prompts are primary; enrichment is opt-in.
(2) Low cognitive buy-in: micro-tasks only (seconds, not minutes), with clear exits.
(3) Context fit: content ties to the drive (e.g., local context, vigilance cues) rather than free-form chatter.

Goal order reflects a deliberate emphasis on cognitive safety before contextual richness. We therefore prioritised *low cognitive buy-in* to minimise the risk of inducing active fatigue (inverse of passive fatigue) [24]. This emphasis ensured that any contextual dialogue remained secondary to maintaining attentional safety.

Once triggered, the CA operated fully autonomously, generating content in real time rather than relying on pre-scripted or Wizard-of-Oz responses. We relied on the live API to preserve authenticity and demonstrate how a deployable agent could operate in the vehicle. The feature was integrated into a bespoke iOS application (see Figure2-a), which allowed experimenters to discreetly position the device out of the driver's view and interact via voice alone. The app supported pre-defined triggers based on time and GPS location, and enabled the input of participant names to deliver more empathetic, personalised dialogue.

The design was refined through approximately 10-15 pilot tests with the broader research team to confirm audio quality, interaction timing, and safety prompts. This allowed us to adjust the balance between naturalness and conciseness, ensuring conversations remained engaging without overloading drivers.

Examples of Zoe's interaction included:

- *Naturalness*: "Nice to meet you, Mindy. I'm Zoe, your driving companion. How are you finding the drive so far? Are you feeling alert and engaged?"
- *Context-awareness*: "Let's keep an eye out together. Have you noticed any wildlife near the track?", "What do you see around us? Anything interesting catching your eye?"

4 Method

A between-subjects test-track study was conducted to investigate the impact of a real-time CA on passive fatigue in the context of conditional automated driving. Perceived usefulness, acceptability, and usability were also assessed via semi-structured interviews. Forty participants each completed a fifty-minute drive in an L3 prototype vehicle on a closed test-track. The chosen duration exceeds that of related studies that employed approximately 25-minute driving periods [17, 23]. Additionally, prior work reports that fatigue (impaired vigilance, increased sleepiness and mind wandering) occurs after as little as 15 minutes of automated driving [47].

Participants were assigned to one of two conditions: CA (received a short, spoken interaction with a real-time agent near the end of the drive) or Control (completed the same drive with no agent interaction). Following the drive, semi-structured interviews were conducted with both groups, during which the control group also experienced the conversational agent (though not in context).

4.1 Participants

Forty participants were recruited (m:f = 24:16, M = 56.8 years, SD = 19.2). Recruitment was conducted via email invitations, social media posts, flyers, and snowballing, targeting diverse communities including campus advertisements, driving clubs, women's community groups, senior centres, and public noticeboards. Participants completed a pre-screening form including demographic details and health-related questions and were provided with study information sufficient to understand potential risks. The sample included younger drivers and older drivers (65+) to examine potential age-related differences in driver states and reactions to technology. All participants held valid driver licences and at least two years of driving experience. Individuals reporting a history of fatigue or sleep-related medical conditions were excluded, and participants were instructed to avoid caffeine and other stimulants or depressants on the day of the session. Participants were allocated to either the CA group (n = 25, m:f = 14:10, M = 53 years, SD = 20.2) or the control group (NCA) (n = 15, m:f = 10:5, M = 55.2 years, SD = 17.2). Given fixed recruitment resources, we adopted an approximately 2:1 allocation favouring CA to maximise data from the condition of interest.

All participants received information about the study, specifically a statement about protecting safety and privacy and signed informed consent and image release forms approved by the Ethics Committee of Queensland University of Technology (approval number 8522). Each participant received AU$150 as compensation for completing the study. This was deemed appropriate due to the participants' on-track time commitment (about 2-3h, including waiting time), and significant travel time to and from the test track (approximately 1.5h).

4.2 Apparatus

4.2.1 Vehicle and SAE Level 3 Automation. The study was conducted using a Renault Zoe electric prototype vehicle adapted for SAE Level 3 (L3) automated driving (Figure2-b). Conducting the experiment in L3 facilitated exposure to authentic vehicle motion, vibration, road curvature, and naturalistic fatigue onset patterns. This methodological choice meaningfully enhances ecological validity and distinguishes this study from prior CA-focused fatigue mitigation research [17, 18, 23]. The vehicle operated in automated mode on a closed-road circuit. A trained safety driver was seated alongside the participant and could assume control at any time, ensuring safe execution of handover and takeover procedures.

4.2.2 Implementation of the Conversational Agent. The CA was deployed as a mobile phone application, programmed to initiate a 120-second dialogue [23, 39] at the start of the sixth monotonous lap on the track. Audio was delivered through bone-conduction headphones. This enabled participants to hear both the agent and ambient vehicle sounds without interference. Participant responses

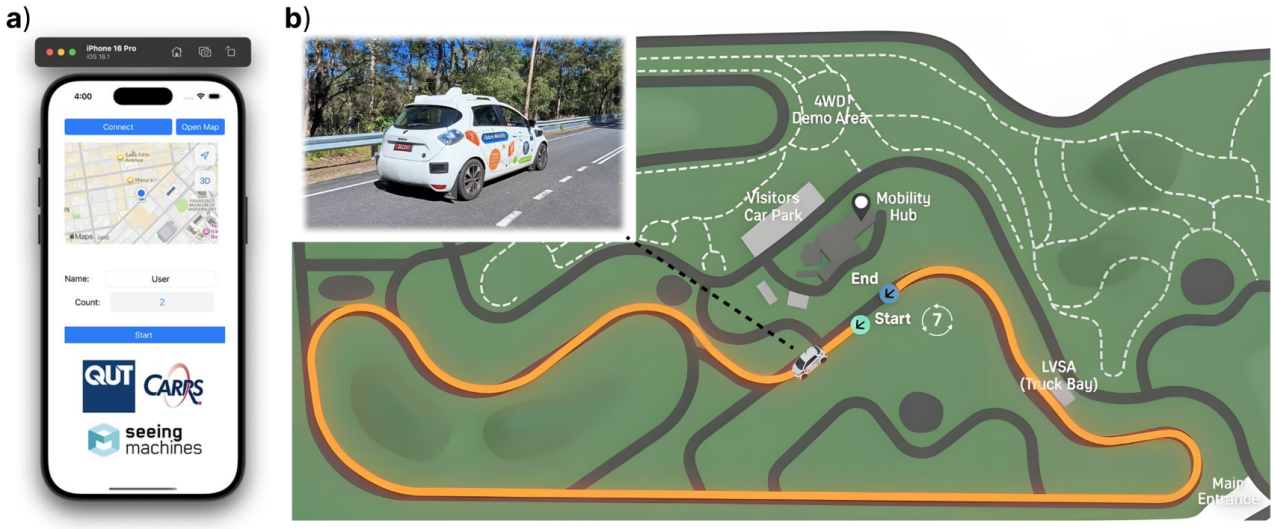


Figure 2: a) The iOS App developed for implementing the CA - Zoe; b) The L3 automated vehicle prototype, Zoe and closed-road circuit employed for monotonous automated driving trials.

were captured via the headphone's built-in microphone, creating a natural, hands-free conversational experience. All conversations were transcribed and stored locally on the device. The prompt development and interaction strategy adopted a LLM-based agent design for combat fatigue based on Yu et al. [50]'s design. For the control (NCA) group, the agent was demonstrated only during the post-drive interview to avoid influencing their in-vehicle experience or introducing carryover effects between conditions. The purpose of this demonstration was to give NCA participants the opportunity to experience the CA after the experiment, allowing them to imagine how this form of interaction might have applied to their own automated drive they had just experienced, and to reflect on this during the post-drive interview (see Appendix Table 2).

4.2.3 Side Video Recording. In-car video was recorded using an iPhone mounted on the safety driver's side window. The camera captured the driver's behaviour, such as posture, attention, and interactions throughout the drive (Figure 1). Recording began after the familiarisation phase and concluded once the participant completed all driving trials. This provided a continuous behavioural record to support analysis of engagement and fatigue.

4.2.4 Karolinska Sleepiness Scale (KSS). The KSS was administered to capture subjective states of passive fatigue during the monotonous driving trial. Widely used in fatigue research [17, 23], the KSS provides a 1–9 scale for participants to self-report their level of vigilance. After the familiarisation phase, the safety driver explained the scale to participants and asked them to report their ratings verbally at predetermined points during the drive (see Section 4.3.3). These were recorded by the safety driver. These scores served as a complementary indicator alongside behavioural observations and interview data. KSS administration may introduce brief attentional resets, potentially mitigating deeper fatigue onset. To minimise this effect, researchers rehearsed KSS ratings twice with participants before the driving trial, and during the drive, KSS ratings were elicited with brief, consistent prompts across conditions.

4.2.5 Semi-structured Interview. Post-drive, participants completed a semi-structured interview with open-ended questions (see Section 4.3.4). Interviews were audio-recorded using a digital pen, with one researcher leading the discussion and another taking field notes. For the NCA group, the agent was demonstrated at this stage by the assisting researcher to elicit comparative feedback.

4.3 Procedure

The study was conducted during daylight hours (08:40–17:00). Each participant completed an on-road test-track session in the AV. The session was structured into four phases: Introduction and Training, Familiarisation, Monotonous Driving, and a Post-Drive Interview (Figure3-a). During the drive, the CA either initiated an interaction with the driver at a predetermined location or remained inactive, resulting in continued monotonous driving.

4.3.1 Part I – Introduction and Training. Participants were welcomed with a briefing and signed informed consent forms. An introductory video presented the study aims broadly as exploring interactions with a state-of-the-art digital CA during automated driving. The video also described system functions and safety protocols, emphasising that an intelligent conversational agent may initiate a short dialogue during the drive. Participants were instructed not to converse with the safety driver once the trial commenced to avoid confounding fatigue measures and the effect of conversation with the CA. This included instructions to not check their phone, watches or interact with the central console. They were also briefed on their responsibilities as conditionally automated drivers, specifically, that they were permitted to remove their hands

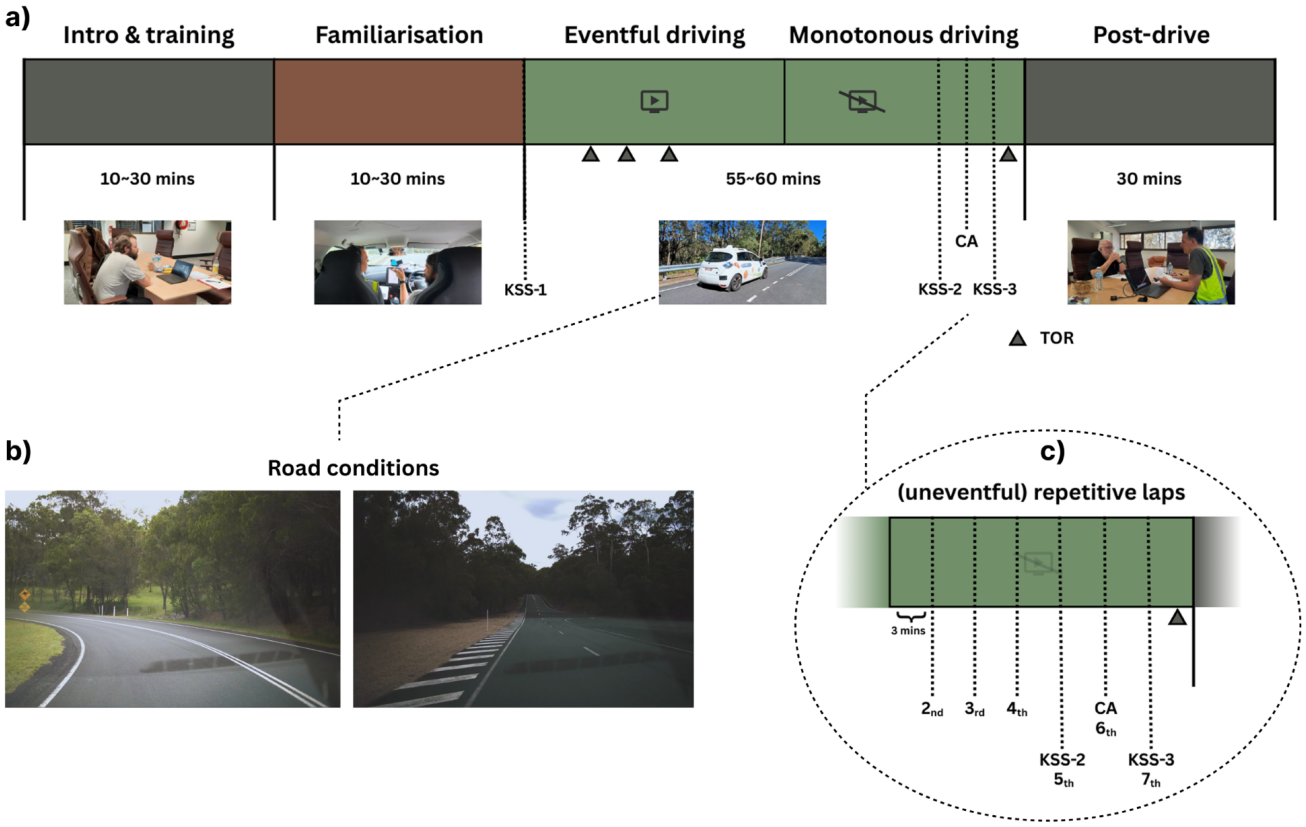


Figure 3: a) Timeline of the experimental protocol, b) the road conditions and c) the details of the monotonous, repetitive laps.

from the steering wheel and engage in limited non-driving-related activities, but were required to remain ready to take over control when prompted. This training included a standardised instructional video that had been used in previous studies for the experimental AV.

4.3.2 Part II – Familiarisation. Participants were guided from the preparation area to the start of the track. To build comfort, trust in the AV and prevent mode misidentification, the safety driver explained the vehicle's functionality and repeated key safety instructions. Participants adjusted the seat and performed a 10-minute warm-up drive, including exposure to automated driving, manual control, and transitions of control (from manual to automated and vice versa).

4.3.3 Part III – Monotonous Driving. Participants then undertook a prolonged driving trial, lasting approximately 55 minutes in total. This driving trial included: 1) Thirty minutes of more eventful automated driving with three planned takeover requests (TORs) while watching videos as an NDRT (all participants watched the same series of short mini-documentaries regardless of condition) and taking different laps and turns around the track. TORs were provided to contextualise the interaction with the CA with respect to mitigating passive fatigue in the context of L3 automated driving, where drivers may be requested to take over. This contextualised the CA as a potential safety feature in the post-drive interview. 2) Followed and contrasted by multiple repeating laps, specifically designed to be passively fatiguing by being uneventful, low-stimulation and repetitive; and 3) A takeover request in the last lap, further emphasising the conditional automated driving context for the post-drive interview. Each lap included straight segments, wide curves, and sharp turns. The vehicle operated in automated mode at a maximum speed of 50 km/h, as required by local regulations, on a single carriageway with two lanes (Figure3-b).

In this session, participants first engaged in a non-driving-related task (watching video clips). The video was then turned off by the safety driver, leaving participants idle and without purposeful activity. This was followed by seven repetitive laps of a closed, low-stimulation circuit (each 3 minutes), designed to induce passive fatigue through monotony (Figure3-c). During the penultimate lap, the CA initiated a 120-second dialogue [23] at a fixed location on the track, prompting drivers to observe and verbally comment on surrounding road and environmental features. Participants in the control condition completed the same trial but without CA activated. A final takeover request due to an emergent lane-changing event was issued during the last lap to re-contextualise the CA in the context of conditional automated driving.

The KSS was administered for driving at the following points:

- Pre-drive (baseline),
- Prior to the activation of the CA (on the start of the 5th monotonous lap, across both conditions),
- Following the CA interaction (on the start of the 7th monotonous lap, across both conditions).

4.3.4 Part IV – Post-Drive. After the drive, participants took part in semi-structured interviews (15–30 minutes) in a quiet preparation room. The interviews explored their subjective experiences of automated driving with or without the CA and probed perceptions of alertness, naturalness, and engagement. Participants in the control group were also given the opportunity to experience the CA, although not within the driving context.

The interview design followed a structured progression:

- *Warm-up* – Participants answered introductory questions about their background and general driving experience.
- *Initial reflections* - Participants were first invited to share their overall reflections on the CA (Q1). Participants is in the control group completed a demonstration with the CA at this point.
- *Disclosure and perceived effects on alertness* - The researcher clarified the design intent of the CA. Participants discussed whether and how the agent influenced their state of alertness, providing specific examples of situations where it was helpful or unhelpful (Q2&3).
- *Interaction experience and enjoyment* - Participants described the naturalness and challenges of conversing with the agent, as well as its entertainment value and impact on enjoyment of the drive (Q4&5).
- *Willingness to use and improvements* - Finally, participants reflected on their likelihood of using such a system in the future and suggested potential design enhancements (Q6&7).

All interviews were audio-recorded with consent and transcribed verbatim. Observational fieldnotes were collected during and after each session to capture behavioural responses and contextual details.

4.4 Analysis

The analysis process comprised three interconnected stages: (1) video analysis of in-car recordings, (2) quantitative assessment of KSS ratings, and (3) thematic analysis of post-drive interviews. These stages were not conducted in isolation; rather, each iteratively informed the others to strengthen interpretation. This approach adapted a concurrent nested analysis, privileging qualitative analysis [44] while using quantitative measures in a complementary role to support interpretation.

4.4.1 Video Analysis of In-Car Recordings. To complement the post-interviews, in-car video recordings were examined to capture observable indicators of attention and driver states that may not have been explicitly reported by participants. The analytical focus was on identifying the types of behaviours present within the observation windows (i.e., during CA interaction or the NCA periods), rather than quantifying their frequency, acknowledging that some behaviours may vary across individuals. Audio-visual data analysis is increasingly employed in HCI and CSCW research to surface 'seen-but-unnoticed' aspects of interaction [6, 12]. Following this, we reviewed 40 in-car side video recordings (each approximately 55 minutes). Three participants were excluded as the video was cut due to the recording phone overheating. Consequently, the analysis focused on 37 segments (approximately 8 minutes each) where the CA was active and compares them to corresponding periods without agent interaction.

For each instance, the analysis centred on fixed track locations where the agent was triggered. Observations were coded within a three-minute window before the conversational onset, a three-minute window during the intervention lap, and a one-and-a-half-minute window after the intervention lap (before the planned takeover, which is out of the scope of this paper). This window was extended when notable behavioural changes persisted. Any behaviours prompted during safety driver-initiated KSS probes were excluded from coding.

Particular attention was paid to body movements (e.g., gestures of thinking, communicating, or showing fatigue), facial expressions (coded using an emotional wheel), and head or eye behaviour (e.g., gaze direction, noticeable changes in blink rate). These cues provide insights into situation awareness [9], arousal [26], and potential states of passive fatigue [6].

To support collaboration, coding was documented in a shared Google Sheet, enabling authors to refine categories and resolve discrepancies. However, all video playback was conducted locally on a secure data drive to maintain participant privacy and data protection. The resulting patterns or insights were then cross-referenced with interview data to explore not only what participants said about their experiences, but also how their behaviour unfolded in situ.

4.4.2 Karolinska Sleepiness Scale. Concurrently, participants' subjective sleepiness ratings on the KSS were analysed to provide quantitative support for the findings. Descriptive and statistical analyses were conducted to examine user differences in KSS scores. Two aspects were of particular interest: (1) whether individuals' KSS scores changed before and after interacting with (or without) the CA, suggesting the agent's potential to mitigate fatigue; and (2) whether these changes aligned with behavioural patterns identified in interviews and video analysis, indicating consistency between subjective reports and observed behaviour. In this way, KSS results offered an additional lens to triangulate and substantiate the themes.

4.4.3 Thematic Analysis of Post-Interview. The interviews were analysed using notebook Qualitative Content Analysis (QCA) procedures [16]. Transcripts were generated with the support of Feishu Minutes and the Whisper API, then manually verified word-for-word by two authors. In total, 39 transcripts were analysed; one participant was excluded for not consenting to the use of AI-enabled transcription tools. The two authors first familiarised themselves with the transcripts before applying a deductive coding approach. Prior to coding, relevant theoretical perspectives for passive fatigue were discussed. This includes the the Yerkes–Dodson law of the relationship between cognitive load and performance [26] and various technology acceptance models [8, 14, 46].

All transcripts were coded in MAXQDA [3]. Codes captured participants' reflections on their experience, attitudes, interactions and

[3] https://www.maxqda.com/

recommendations with both the automated driving system and the CA. To ensure rigour, Authors 1 and 2 independently coded an initial 20% of the data. Inter-coder reliability was calculated and indicated strong agreement (Cohenś k = 0.84). Discrepancies were resolved through discussion and iteration before proceeding.

In subsequent rounds, the remaining interviews were divided between the two authors, with regular meetings to iteratively refine and consolidate the codebook. Emerging codes were organised into initial themes, which were further discussed with the wider research team. Insights from video analysis and KSS results were later used to challenge, support, and deepen the thematic interpretations.

Two authors jointly conducted the video coding, KSS analysis, and thematic interpretation. These analyses were carried out in parallel, with thematic analysis iteratively informed by insights from the video and KSS results. Overall, the video data revealed how drivers behaved when interacting with the CA, the KSS captured their subjective states before and after the intervention, and the thematic analysis investigated perceptions of usage and effectiveness through participants' accounts. Together, these sources of evidence were triangulated to refine and interpret the findings.

5 Results

5.1 Drivers' Behaviour while Driving

Figure 4 summarises the common behaviours observed, along with any unexpected driving behaviours.

Fatigue-related behaviours During the observational window prior to the intervention, participants frequently displayed behaviours linked to drowsiness. These included yawning, slow blinking, and partial eye-lid closure. Instances of wide-open eyes or frowning also indirectly indicated attempts to fight sleepiness. Following the CA intervention (CA group), some instances of yawning and slow blinking persisted.

Small, restless movements were common and interpreted as signs of boredom. Examples included looking at nearby objects (such as the side door, a watch or the centre console), and oral activities like licking lips, lip pressing, pouting, self-talk, or whistling. Hand movements were frequent, such as scratching the neck, thighs, or arms; tapping rhythmically on a palm or thigh; twirling fingers; and touching the eyes, face, hair, forehead, or nose. Postural shifts were also observed, including crossed arms, a thinking pose, and leaning on the window. Some combined behaviours, such as adopting a thinking pose while looking upward or fast blinking, suggested that attention was directed away from driving-relevant information. During and after the CA interaction (within the coding windows), these small movements decreased.

Overall, these observations illustrate how drivers behaved in situ during monotonous L3 driving, revealing both drowsiness-related signs and boredom-induced activity.

Relaxation and alert. Both relaxed and alert behaviours were observed, appearing before and after the CA intervention (Figure 5). Relaxed behaviours included stretching the arms, shoulders, or legs; slightly tilting the head upward; leaning against the window; looking around with broad head movements; and adjusting posture in the seat.

Some participants appeared to remain alert, as shown by extended periods of looking straight ahead or scanning the surroundings while maintaining an upright posture. One participant stated during the post-agent KSS probe, "I won't trust the system," to the safety driver. However, relaxation and alertness were not mutually exclusive. Drivers who kept their gaze forward for relatively long periods also occasionally displayed relaxed or fatigued behaviours. In the NCA group, one participant temporarily took control to avoid roadside wildlife but soon reverted to behaviours reflecting tiredness and relaxation.

These observations suggest that relaxation and alertness were dynamic states, potentially shaped by trust in the AV or by driving experience or habit (see thematic section).

Engagement and attraction. Compared with the NCA group, CA participants displayed positive affect during and after conversations (Figure 5). Facial expressions such as smiling, laughing, or showing surprise were observed, and some participants joked with or were amused by the agent. Following the framework of Plutchik [34], these expressions were coded as Joy and Ecstasy. Others engaged less, responding only briefly. For example, one participant rejected the CAs opening gambit, replying; "Not really… but thanks for asking, you keep me awake." Down-turned mouth corners were also identified as micro-expressions, potentially signalling dissatisfaction with the content.

Participants increased their scanning of the surroundings and gave verbal responses in response to the agent's prompting. Cognitive activities during conversation were also observed, such as lateral eye movements and thinking pose. Some also used hand gestures while speaking, typically between the steering wheel and the head. One participant instinctively touched the steering wheel when the voice was first activated, then relaxed and engaged once realising it was not a takeover request.

Participants also expressed varied communication needs. Some requested functional features, such as *"Can you play some music?"*, while others sought entertainment, asking *"Let's play a word game."* After the dialogue ended, a few attempted to re-engage Zoe with prompts like *"AI agent, what's your name?"*. This indicates a desire to continue, though they were unaware that the system was programmed to stop after a fixed location.

The above behaviours were not observed in the NCA group. Overall, conversations with the CA elicited positive emotions associated with joyfulness, combined with gaze behaviours and hand movements. Participants expressed willingness to engage with the CA, though their needs and expectations varied widely.

5.2 Self-reported Sleepiness Scale

A Mauchly's test was firstly conducted, indicating that the assumption of sphericity was met, $\chi^2(2) = 0.84, p = .656$. Thus, a 2 (Group: CA vs. NCA) × 3 (Time: Baseline, Pre, Post) mixed-design ANOVA was used to examine participants' KSS ratings. The analysis revealed a significant main effect of *Time*, $F(2, 76) = 9.93, p < .001, \eta^2 = .21$, and a significant *Time* × *Group* interaction, $F(2, 76) = 5.51, p = .006, \eta^2 = .13$. The main effect of *Group* was not significant, $F(1, 38) = 0.32, p = .575, \eta^2 = .01$.



Figure 4: In-situ fatigue-related behaviour examples from participants in L3 driving context.



Figure 5: Examples of passive, relaxed states and active, engaged behaviours during the monotonous drive and when interacting with the conversational agent.

The results indicate that participants' KSS scores overall differed across the three *Time* points. In general, the KSS score increased from Baseline ($M = 3.15, SD = 0.24$, $p < .001$) to Pre ($M = 4.29, SD = 0.32$) and decreased from Pre to Post ($M = 4.03, SD = 0.28$, $p = .003$). When considering the interaction between Time and Group, the results revealed divergent trajectories (Figure 6). For the group with CA, the KSS measured in Pre was significantly higher than in Baseline ($p < .001$) and Post ($p = .002$). For the NCA group, the KSS in Post was significantly higher than the Baseline ($p = .004$). The difference between CA and NCA in KSS was only significant in Post ($p = .028$).

These results indicate that passive fatigue increased over time in the absence of an intervention (NCA). However, engaging with the CA appeared to reverse this trend. At the two earlier time points, before the onset of L3 automated driving (Baseline) and before the interaction phase (Pre-agent), the CA and NCA groups did not differ significantly in drivers' levels of alertness. In contrast, following the intervention phase (Post-agent), the alertness levels of drivers in the CA and NCA groups diverged. As hypothesised, participants reported increased alertness following the CA interaction.

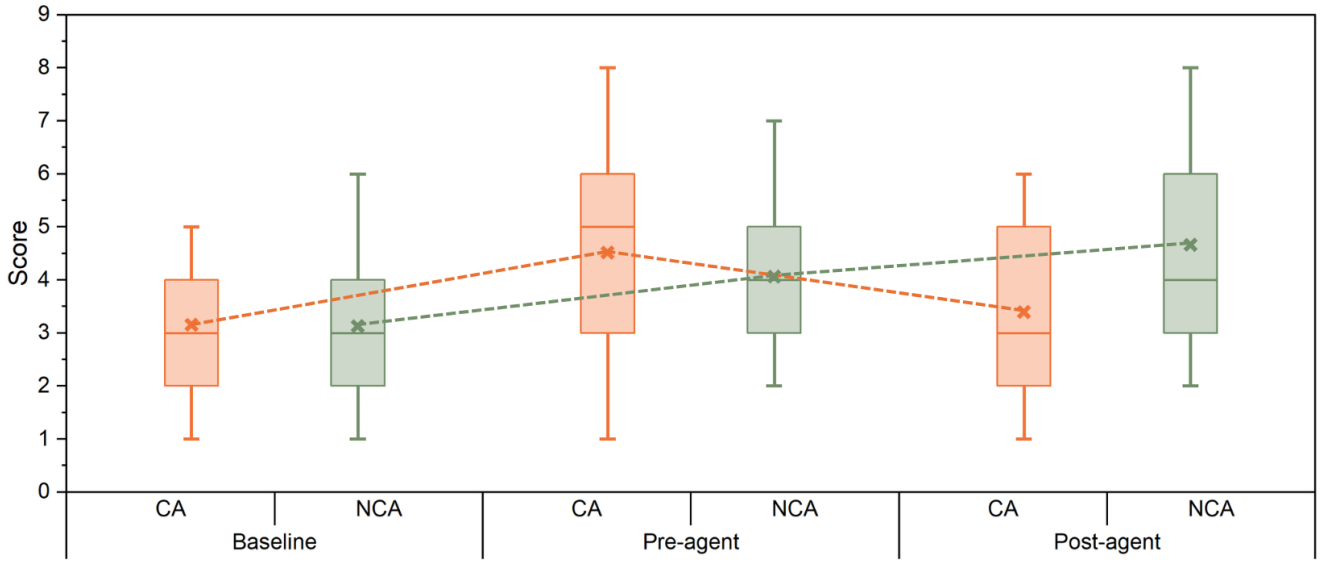


Figure 6: Karolinska Sleepiness Scale (KSS) scores at baseline, pre-interaction, and post-interaction.

5.3 Thematic Analysis

Thematic analysis was used to 1) answer research questions regarding user perceptions of the CA as a useful safety tool, and 2) factors that affect the effective design of a safety-oriented CA for conditionally automated driving. The following sections will describe findings relating to the user experiences of both of these aspects.

Theme 1. Perceptions of Usefulness for Managing Passive Fatigue. Extending on observations and measures of vigilance, we explored how drivers perceived the CA as influencing their alertness. Three distinct themes within the overall perceived usefulness of the CA emerged: (1) interrupting sleep-related processes, (2) interrupting boredom-related processes, and (3) supporting alertness and engagement. These align with the core symptoms of passive fatigue presented by [24].

Interrupting sleep-related processes Participants described the CA as useful and effective for interrupting sleep-related processes during automated driving. This was consistent across age, gender, condition and perceptions of usability of the agent. This finding corroborated with KSS scores, suggesting the CA was useful for waking participants up. Quotes of the following sentiment were nearly unanimous across participants.

> *"Definitely felt a level of awakeness, kick in, which is good."* -P12

> *"I felt sleepy, then this voice may alert me again when driving."* -P02

These quotes effectively demonstrate the potential of the agent to interrupt sleep-related processes. Thus some drivers found that the agent useful for interrupting sleepiness but still found the interaction boring or distracting.

Interrupting boredom-related processes Some drivers saw the agent as breaking monotony or adding entertainment, others found it dull or superficial. Drivers who found the agent entertaining were often excited about the potential of such an agent in the future.

> *"I think it would be positive, um, because... it's asking you questions and you're responding, it's sort of getting you to refocus and start using your brain differently again."* -P95

These users were also likely to see great potential in future iterations of similar agents:

> *"So it's certainly entertaining. And I think potential is boundless, it's just really cool".* -P01

In contrast, some drivers found the interaction to be boring and unlikely to keep them engaged with the agent for an extended period of time. This was unrelated to their perceptions of the agent as a safety unit.

> *"It was boring. It was... alerting, but, yeah, really boring."* -P04

> *"That's not what I call entertainment."* -P08

While many participants offered suggestions for improving the entertainment value of the system, some stated that they would neither want nor need entertainment from such an agent. This was often accompanied with an explanation that human–agent conversation in any form simply did not align with their preferred forms of in-car entertainment.

> *"I'm sure they [other users] would [be entertained by the CA]. I'm sure some people would have lots of fun with it, really."* -P80

> *"That would fit some people. I don't know I really enjoy that."* -P49

However, these users who did not enjoy the gamified aspects of the interaction did accept the agent as a necessary safety tool for AV driving.

"I wouldn't want to talk politics or anything, but yeah, just as an alertness thing now every, so if you do a long hour long drive, maybe every 20 minutes." -P99

Supporting Alertness and Engagement The data also revealed strong support for the CA as an alertness tool that actively re-engaged drivers with their surroundings.

"It made me refocus. ... it reminded by asking about the signs, it sort of gets you to focus back on the, on the narrow scope of the road." -P03

This was also framed as providing a passenger-like presence that brought drivers back 'in-the-loop'.

"It's kind of like having that… you bring a passenger along… less jarring than an auditory tone alert." -P94

"It actually made me pretty alert… switched from a bit dizzy mode to a proactive mode because I have started conversation with a person." -P08

However, some drivers conveyed that the agent limited their ability to focus on the road. This idea was more common in drivers who did not respond to the fatigue intervention (KSS scores did not increase from Baseline to Pre-agent). Distraction was one of the most common suggestions among *non-responders*:

"I feel like it [the CA] *could be a good thing or also a bad thing at the same time, because when you're driving, you try to make sure that you don't have a lot of distraction."* -P31 (NCA)

"A distraction ... until you got used to it. And then asking, what do you see around you? Some people might take their eyes off the road, literally, to have a look around. I think that might be a bit of a distraction. -P79 (NCA)

Even when participants were immersed in prolonged conditional automated driving, their responses were often framed in relation to manual driving rather than hypothetical future automated driving. This suggests that perceptions of safety-oriented CAs may be affected by mode awareness. Such awareness can influence the extent of driver engagement and, in some cases, lead the CA to be perceived as a potential distraction.

"I already know what I'd do to help me stay engaged on a long drive, but then again, it wasn't automated, so I honestly don't know. So I normally just put on a podcast or music or something like that, and that's enough to keep me engaged." -P12

"Yeah, it can be great ... but, if I'm in the manual mode and I have to concentrate on a lot of stuff on the road, sometimes it can be distracting. " -P25

In conclusion, participants perceived the CA as useful in an AV context for reducing fatigue. However, these effects appeared to be shaped by multiple factors, including the content of the CA's dialogue, individual characteristics such as personality and driving experience, the degree of agent type alignment with user needs, and contextual influences such as mode awareness.

Theme 2. Design and Interaction Factors

This theme summarises the factors identified by participants as influencing the adoption of a CA. These factors centred on three key areas: user acceptance, accessibility, and usability. Each requires careful consideration by designers to ensure the system is appropriately adapted for conditional automated driving safety scenarios.

Perceived Benefits Underpinning Acceptance Despite unanimous agreement that the agent was helpful as a safety tool and increased alertness, acceptance varied considerably in terms of whether participants would want such a safety feature in their own vehicles. Some were enthusiastic despite the technological limitations associated with the experimental nature of the agent's capability at the time of the study. Reported benefits included entertainment (e.g., playing games with the agent), safety (e.g., drawing attention or seeking a coffee replacement), and companionship (e.g., reducing loneliness).

"I didn't like the game that we played, right? But yeah, the concept of it popping in and being available to talk through different things or play different games for that kind of thing is really good." -P16

"It's going to sort of make me look about a bit more, take notice more of what's occurring... you know, the direction we're going in or signs or whatever... I didn't have any issues with that." -P89

"It would be quite beneficial, I think, to have something like that if you were just travelling alone," - P80 (NCA)

By contrast, others emphatically rejected the system for a variety of reasons. Prominent concerns were that some drivers simply disliked conversing with a machine.

"I just find it weird engaging with an LLM… you don't really get anything out of it." -P12

"With the conversational agent, yeah, I don't know. I just found her like a computer box... which I found annoying". -P84

A distinct group of participants also emerged who suggested they accepted the agent but only with certain provisions and only as a safety tool.

"Just as an alertness thing… every 20 minutes, two or three minutes, nothing more." -P99

"Although I think that people will get annoyed very quickly with these things, so that would have to be, you know, in the regulations saying, like, they've got to check you every 10 minutes and you've got to respond in your voice on some sort of level or scales." -P04

Even among users who found the agent helpful for supporting their engagement with the driving task, acceptance pended on increased personalisation in future CAs. This idea applied to design cosmetics and presentation of the agent.

"It's something I'd consider, but it would depend on, what the options with the car. Things like voice, I like to be able to customise to a different voice." -P16

Another strong theme was whether the agent was able to accommodate the users' unique desires, whims and nuances:

"I think I would be more alert if it would be responsive to my desires to talk about some sort of a topic, not just being alert because you need to pick up some more wildlife, so what do you see? ... However, I found sort of a limit where it wants to stay in this alertness domain, right? So it doesn't want to talk about what's going on in US politics, right? It wants to stay there." -P01

"You need a better AI than that... insufficient response to, I mean, it has a series of program responses, but the program is very small. You're not interacting with me in the true sense... I'm assuming

that it would be programmable, this thing... you know, you have one that tells you a story, one that tells you a book." - P113

These quotes illustrate one of the strongest sentiments presented across users; that consistent future use would pend on the ability of the CA to shape to their preferences and accommodate their preferred engagement. This ranged from *minimalist*, safety-oriented users who would prefer to only be disturbed for safety-critical reasons, users who preference all-in-one entertainment and integration to social users who desire meaningful and personal connection.

Accessibility-Related Barriers

Although accessibility was not an initial focus, interview themes highlighted the need for CAs to be adaptable. Unlike prior simulator-based studies [13, 17], participants in the real-world setting not only discussed communication frequency and complexity, but also raised higher-level requirements affecting engagement—such as dialect, accent, and hearing-friendly dialogue.

"Didn't always pick up what I said… maybe my pronunciation." -P01

"It's quite good because she's clear... She has a clear, sharp voice, good diction, and she doesn't speak too quickly... And in the world around me, people speak so quickly. My ears are 80 years old." -P113

Usability-Related Challenges

Participants also raised a range of usability concerns that may have shaped their willingness to adopt a CA. As expected, technical limitations, latency, and clunky error recovery were barriers to smooth use.

"I will profess, I don't use Siri or anything else like that. So I also don't use ChatGPT either, so the whole concept of that sort of thing is just random." -P03

Perceptions of the interaction's *naturalness* also varied across participants, likely reflecting differences in user expectations and prior experience with artificial intelligence.

"People would like it, I think. How it could do things better used to be, I felt, at least I had the impression, that it sort of was programmed. It was sort of, you know, like pre-recorded track. It wasn't really engaging and following to what I'm saying to him." -P01

"The whole concept of talking to an AI thing, I thought, was really unnatural." -P03

Interestingly, this sentiment was not shared by all participants. Many perceived that the interaction was in fact natural and referenced feeling like the interaction was shared with another real person.

"It feels like you have someone to tap your back when you are excited. You need to be awake, man. Something like that... This one feels a bit organic." -P31 (NCA)

"It makes it more fun, I would say it was fun, like talking to someone, or it's pretty much like you're talking to him, and sitting next to you, so yeah, it's pretty much like that." -P08

Individual differences in cognitive load Another key finding was that usability was shaped by participants' expectations of how well the CA could regulate workload. Some perceived the conversation as too shallow to sustain alertness (*underload*), while others experienced it as effortful or intrusive (*overload*). Balancing these divergent needs emerged as a central challenge for CA design.

Underload: *"I think in the short term, it would be good, but I mean, how long can you maintain a surface-level conversation with the robot and still be engaged?"* -P12

Overload: *"No one wants to carry a conversation while they're driving, because that's just increased effort."* -P12

Overall, beyond the technological limitations already noted as affecting usability, participants' perceptions of naturalness in real driving contexts were far from uniform. Some held very high expectations for natural conversational exchange, while others valued the agent more as meaningful companionship. In addition, participants' feedback emphasised that a CA should be able to adapt precisely to different users' workload, striking a balance between underload and overload.

6 Discussion

6.1 Emerging Archetypes of Conversational Agent Use

A central contribution of this work is the demonstration that drivers recognised the CA as a useful safety feature. Across conditions and age groups, participants consistently described the CA as helping to wake drivers up. This baseline provides strong evidence that CAs are perceived as legitimate safety tools, not simply novel entertainment for L3 driving. Taken together, these findings extend prior simulator-based work [17, 18, 23] by demonstrating that conversational dialogue can function as an authentic safety tool in a real L3 vehicle with a live LLM-based agent. This may provide novel empirical evidence of how real-time HCI interactions are experienced when embedded within an actual driving system.

Our multi-stage analysis (interviews, in-car video, and KSS) revealed distinct user preference profiles, which we interpret as archetypes. These profiles emerged inductively from qualitative content analysis and align with established CA archetype frameworks (Service/Productivity, Entertainment, Companion) [21]. As participants repeatedly indicated, the value of the CA depended on whether it aligned with their own orientation to the driving experience. Some drivers welcomed short safety-focused prompts, others expected entertainment or novelty to relieve monotony, and still others valued the CA as a social presence. These divergent perspectives point to the existence of distinct user archetypes, patterns of preference and orientation that shaped not only perceptions of usefulness but also future intention to use. This is crucial to support continued acceptance of a CA system as mismatched user-agent needs affect intention-to-use [46].

Although our priority was to design a safety agent whose job was strictly to discuss the immediate driving environment, users explained they would simply turn off a device that continued to engage them in discussions that were not of interest/priority for them. Thus, accommodating these diverse user-needs by presenting *personas* can be seen as key to promoting continued engagement with CAs in automated driving.

Beyond simulator-based findings in this domain, our on-road results show that a brief, natural exchange can function as a safety

intervention without resorting to intrusive alarms, a design approach participants frequently described as "bringing [them] back to alertness." This strengthens the claim that LLM-based conversational dialogue can operate as a gentle but effective attention management strategy in authentic L3 conditions, rather than as novel NDRT entertainment only.

Identification of Archetypes. The heterogeneity in responses can be understood through the Technology Acceptance Model 3 (TAM3) [46]. While perceived usefulness was consistently acknowledged, it was insufficient to guarantee intention to use unless participants also experienced the CA as *job relevant*. For safety-first drivers, the system's narrow focus on road-related prompts matched their priorities: *"To talk about the environment or about the road or about speed or the weather, something which you need to be noting anyway"* (P104). For these users, usefulness and relevance were tightly coupled, supporting adoption.

By contrast, other participants experienced a mismatch between the agent's safety-first framing and their own needs. They tend to prioritise hedonistic needs. Entertainment-seeking drivers valued stimulation over narrow functionality: *"It could be enjoyable… if the automated system can introduce different things, rather than just speaking. You can play music or something, or there is a visual screen… if you can, like, change from time to time the experience… that would be great as well"* (P25). Socially oriented drivers, meanwhile, emphasised companionship and empathy: *"It was good. It was really helpful. I thought that, yeah, someone else is here to help me… it was feeling so good"* (P39). For these groups, safety benefits were recognised but did not align with their self-defined goals, leading to reluctance despite positive effects on alertness.

This mismatch is critical. Without job relevance, a system risks being perceived as intrusive or irritating even when drivers accept its potential benefits. In the context of TAM3, the absence of alignment between agent type and user orientation threatens future intention to use. Thus, safeguarding adoption requires recognising and addressing variation in user archetypes. This contrasts our design of a safety-priority unit which was hypothesised to appeal to the most amount of participants.

Here, our analysis draws on the work of Nielsen [32], emphasising the identification of patterns of user difference before fictionalising characters. Through thematic analysis of interviews, triangulated with behavioural and KSS data, we derived three archetypes: **Safety-First**, **Entertainment-Seeking**, and **Social-Connection Oriented** (see Table 1). To further conceptualise their implications, we positioned them against the design archetypes articulated by Lessio and Morris [21]: Service/Productivity, Entertainment, and Companion. This alignment underscores that not all drivers benefit from a single CA type. Instead, each archetype corresponds to a distinct design orientation with its own criteria for acceptance. Simultaneously, the data shows that traits and roles are not exclusive; real users (and real agents) can exhibit overlapping characteristics, which argues for designs that acknowledge fluidity rather than enforce singular types. Indeed, users displayed shifting needs and expectations throughout the drive.

Table 1: Archetypes of Conversational Agent Users

Safety-First Users	
Trait profile	High monotony tolerance; very high safety concern; high alertness; negligible hedonic motivation; very low social need.
Motivation	To remain alert and maintain vigilance during automated driving; any interaction must clearly serve a safety-related purpose.
Acceptance Interpretation	Perceived usefulness is the strongest determinant of intention to use. Continued engagement depends on whether the conversational agent clearly contributes to reduced fatigue and improved readiness.
Illustrative quotes	"To talk about the environment or about the road or about speed or the weather, something which you need to be noting anyway. The problem is people with passengers in the car, they often talk about things that are totally unrelated." -P104
Entertainment-Seekers	
Trait profile	Very high hedonic motivation; moderate social need; moderate safety concern; low alertness; low monotony tolerance.
Motivation	To combat boredom and make driving more stimulating through novelty, humour, or game-like interactions.
Acceptance Interpretation	Perceived enjoyment dominates intention to use. Safety benefits are accepted as secondary outcomes of staying mentally stimulated.
Illustrative quotes	"I think it's enjoyable, but not most. But it's enjoyable. It could be enjoyable… if the automated system can introduce different things, rather than just speaking. You can play music or something, or there is a visual screen… if you can, like, change from time to time the experience given by the automated system, that would be great as well." -P25
Social-Connection Oriented Users	
Trait profile	Very high social need; moderate hedonic value; moderate safety concern; moderate alertness; moderate monotony tolerance.
Motivation	To experience companionship and empathic interaction during monotonous driving; relational quality is central.
Acceptance Interpretation	Intention to use is driven by a combination of usefulness, enjoyment, and social influence. The system must feel engaging and socially rewarding while still contributing to alertness.
Illustrative quotes	"It was good. It was really helpful. I thought that, yeah, someone else is here to help me… it was feeling so good." -P39

6.2 Linking Driver Archetypes to Conversational Agent Design Archetypes

Lessio and Morris [21] propose five CA design archetypes: *Service, Companion, Entertainment, Care,* and *Productivity*. We compared our three driver archetypes with this framework:

- **Safety-First Users → Service & Productive Archetype** Safety-first drivers align most closely with the *Service* archetype, which is designed to provide clear, reliable, and functional support. Like service-oriented CAs (e.g., Alexa in its productivity focused mode), these drivers value task alignment, precision, and transparency over relational or hedonic qualities.
- **Entertainment-Seekers → Entertainment Archetype** Entertainment-driven drivers map directly to the *Entertainment* archetype. These CAs emphasise play, humour, and novelty, sustaining engagement by offering stimulation. The fit is particularly strong with Nielsen's finding that perceived enjoyment predicts continued use in this group [32].
- **Social-Connection Oriented Users → Companion Archetype** Socially motivated drivers correspond to the *Companion* archetype. Companion CAs, such as Replika[4] or XiaoIce[5], prioritise empathy, memory, and continuity of relationship. For these drivers, relational depth is the primary design quality that sustains acceptance and engagement.

By systematically identifying archetypes and mapping them to established CA design archetypes, we show that not all drivers

[4]https://replika.com
[5]https://www.xiaoice.com

benefit from CAs in the same way. Instead, *safety-first, entertainment, and social-connection oriented users* align with distinct CA archetypes—Service, Entertainment, and Companion, respectively. This mapping provides both empirical grounding (through participant accounts) and future design heuristics (by connecting to a recognised archetype framework). Crucially, tailoring CA design to these archetypes with *agent personas* may increase *intention to use* as theorised by the acceptance model [46], thereby ensuring not only engagement but also improved safety outcomes in semi-automated driving contexts.

6.3 Effectively Design CAs for Attention Management

CA for richer capabilities. Returning to our findings, driver responses, subjective reports, and interviews consistently suggest that CAs can help reduce passive fatigue in an authentic L3 driving context. KSS ratings showed opposite trends before and after interaction with the CA compared to the control group. This aligns with prior simulator-based results [17].

Participants also pointed out performance limitations. Similar constraints were noted in recent simulator studies, which found that low-complexity dialogue was most effective [13]. While this outcome seems intuitive, prior work rarely engaged with a broad set of users in situ to discuss deeper design trade-offs. Our study also introduced environment-related prompts that supported driver alertness, reducing the attention gap between automated and manual driving [20]. Although this paper did not include eye-tracking data, video analysis showed more searching behaviour after CA use, whereas control participants showed no similar change.

Positive behaviours were also observed. Some drivers laughed, joked, or otherwise displayed enjoyment when interacting with the CA. These reactions point to the agent's potential to raise arousal through playfulness. Cognitive activities were also noted, such as thoughtful expressions, suggesting that the CA could increase cognitive load in monotonous driving contexts. A few even gestured towards the CA, unintentionally bringing their hands closer to the wheel and increasing physical readiness. Compared to monotonous NDRTs previously used to mitigate passive fatigue [30], the CA offered richer capabilities.

Safety cues: explicit or implicit? The overarching goal remains to provide effective guidance: preventing underload while supporting driver alertness to facilitate safe L3 driving, and doing so with smooth transitions that preserve user experience [27]. The challenge is how to embed safety functions into CA interactions. User engagement, as discussed in Section 6.1, is a prerequisite: drivers should be willing to use the system and not find it irritating. But should safety cues be made explicit or remain implicit? Our findings highlight three considerations:

(1) **Mode awareness.** Some participants argued that talking while driving was unsafe, even conflicting with road safety campaigns. This shows the importance of situating CAs in the right contexts (i.e. right modes) and building accurate mental models of their use.
(2) **Trust in automation.** Several participants, already in a heightened state of alertness due to distrust of AVs, found the CA distracting rather than helpful. Adaptive timing of interventions, possibly informed by a driver monitoring system, is therefore essential.
(3) **Conversation topics.** A few participants expressed interest in deeper interaction, such as playing puzzle games to "get the brain working." While promising, overly complex dialogue risks cognitive overload and degraded performance [26]. Future LLM-based CAs should therefore integrate control strategies to balance the user requests for in-depth conversation. Designers can manipulate (i) gate conversational depth by workload cues; (ii) offer archetype switching via quick controls (e.g., "Safety-only", "Safety+Entertainment", "Safety+Social"); and (iii) anchor enrichment (games, music, local info, small talk) to explicit user preference and driver state.

In summary, our study surfaces design considerations critical for balancing safety and user experience in authentic automated driving contexts. Yet open questions remain. For example, how long can conversation effects persist before drivers revert to fatigue? What alternative or escalated strategies are effective when the agent fails to meet user expectations? These questions extend beyond the scope of this paper but represent important avenues for future research.

7 Limitations

This study presents early-stage insights based on a prototype AV platform and an LLM-based CA deployed in a dynamic field setting. While the on-road environment adds ecological validity, both the vehicle and the agent were prototypes. The AV operated on a rule- and map-based platform rather than production-level autonomy, and the agent occasionally experienced network-related delays due to the setting. The agent latency fluctuated, which may have provided an inconsistent experience between users. However, this is comparable to standard online voice-based LLM applications, and most participants still rated interactions as natural and responsive.

Participants undertook a single 55-minute session, meaning long-term adoption, novelty effects, or habituation remain unexplored. Indeed, some participants suggested that their perceptions may change given time and exposure. Whilst the sample was diverse in age, targeted recruitment of older and younger drivers to facilitate usability exploration across the age range may affect fatigue ratings. Specifically, older drivers may be less prone to fatigue manipulations in AVs [1]. Additionally, although no age-related differences in perceptions of usability emerged during qualitative analysis, between-subjects hypothesis testing may reveal distinct differences.

Whilst participant allocation favoured the CA condition, all participants experienced the same vehicle, route, environmental, and procedural conditions. None of the participants reported prior experience with L3 during screening or interviews, and such exposure was therefore not considered a confounding factor. Individual differences in communication style or familiarity with CAs were treated as natural variation relevant to understanding diverse user responses rather than as variables to be experimentally controlled.

We also acknowledge that anthropomorphic and personality cues may shape user trust, engagement, and perceptions of agency

[33, 36]. The agent's gendered name ("Zoe") may likewise have contributed to user perceptions [35], and future work should consider how naming conventions and gender presentation interact with these social dynamics.

Beyond this consensus, our analysis revealed three preference orientations (Safety-First, Social-Connection, and Entertainment-Seeking). We deliberately describe these as archetypes rather than fixed categories. Participants' needs and orientations were not static, and individuals often displayed features of more than one profile or shifted between them during the drive. This fluidity is not a limitation of the finding but an opportunity for design. LLM-powered CAs are uniquely suited to accommodate dynamic shifts in user preferences through adaptive dialogue. Covering the three primary orientations observed here, therefore, provides a practical foundation for designing agents that can meet the majority of user needs while remaining sensitive to contextual changes.

Together, these limitations point to the need for deeper longitudinal, cross-cultural, and adaptive investigations. At the same time, they highlight the value of our study as a proof of concept: demonstrating that even under constrained conditions, conversational interaction can be deployed authentically in a real AV, yielding insights into how users experience, accept, and desire such agents.

8 Conclusion

This study demonstrates the potential of CAs to mitigate passive fatigue in conditional automated driving. Rather than relying on direct safety alerts, the approach embedded safety goals into natural, environment-related dialogue. In a real-world test-track study, users consistently identified the agent as a legitimate, prospective safety tool. Users described how brief conversational exchanges woke them up and re-engaged their attention with the driving environment, helping manage passive fatigue symptoms. Perceptions of the entertainment value of the agent were mixed. Overall, these findings provide strong evidence that LLM-powered conversational dialogue can serve as a credible intervention to support safety in L3 driving contexts.

Our analysis also revealed three predominant user preference profiles, Safety-First, Entertainment-Seeking, and Social-Connection oriented drivers. Aligning these orientations with established CA design archetypes [21] highlights a central HCI contribution. Acceptance is not determined by usefulness alone but by the fit between agent behaviour and drivers' own goals. For some, safety cues were paramount; for others, stimulation or companionship mattered most. Designing for conditional automation, therefore, requires agents that can flexibly accommodate diverse orientations without compromising safety. These insights provide preliminary design heuristics for future passive fatigue management CA development.

Methodologically, the work advances simulator and Wizard-of-Oz approaches [17, 18] to study live LLM-based agents in L3 AVs. The integration of behavioural video, sleepiness ratings, and qualitative interviews offers a reusable model for evaluating conversational interaction in safety-critical domains. This approach demonstrates how early-stage prototypes can still yield valuable insights into user acceptance, experience, and usability evaluations.

Together, these contributions position conversational agents as future adaptive HCI interventions that bridge the gap between passive supervision and active readiness. By systematically linking empirical archetypes to design archetype frameworks and acceptance models, the paper provides conceptual and practical guidance for designing usable and desirable CAs in future AVs. This alignment offers practical guidance for tailoring agent personalities and interaction strategies to diverse driver needs, while recognising that user orientations may shift dynamically over time. These insights can be generalised to other safety-critical supervisory contexts, from aviation to rail and control rooms, where cognitive underload continues to threaten vigilance. By offering proof-of-concept and design guidance for fatigue-intervention conversational agents, we show how small talk can yield big impacts for the future of human–machine teaming.

Acknowledgments

This research was undertaken under the QUT Industry Chair in Empathic Machines, in collaboration with and co-funded by industry partner Seeing Machines Ltd. Through participation in the National Industry PhD Program (NIPP), it has also been co-funded by the Department of Education. The Australian Research Council also provided financial support for this project through the Australian Research Council Industrial Transformation Training Centre for Automated Vehicles in Rural and Remote Regions (IC230100001).

We thank the Queensland Department of Transport and Main Roads for providing access to the highly automated prototype vehicle Zoe2. We also thank the entire research team for their contributions, in particular the safety drivers who ensured operational safety throughout the study. We further acknowledge Sean Baek from the Faculty of Science for his support in software development, the research assistants who supported data collection, and the staff at the RACQ Mobility Centre for their support and understanding during the on-track data collection.

References

[1] Sadegh Arefnezhad, Arno Eichberger, and Ioana Victoria Koglbauer. 2022. Effects of automation and fatigue on drivers from various age groups. *Safety (Basel)* 8, 2 (April 2022), 30. http://doi.org/10.3390/safety8020030

[2] Carryl L Baldwin and Ian McCandliss. 2019. Attention management in highly autonomous driving. In *Human Performance in Automated and Autonomous Systems* (1st edition ed.). CRC Press, Boca Raton, FL : CRC Press/Taylor & Francis Group, 2019., 231–248. https://doi.org/10.1201/9780429458330

[3] Lukas Bier, Michael Emele, Kaja Gut, Jasna Kulenovic, David Rzany, Max Peter, and Bettina Abendroth. 2019. Preventing the risks of monotony related fatigue while driving through gamification. *Eur. Transp. Res. Rev.* 11, 1 (Dec. 2019), 1–19. http://doi.org/10.1186/s12544-019-0382-4

[4] Cristiana Bolchini, Carlo A Curino, Elisa Quintarelli, Fabio A Schreiber, and Letizia Tanca. 2007. A data-oriented survey of context models. *SIGMOD Rec.* 36, 4 (Dec. 2007), 19–26. http://doi.org/10.1145/1361348.1361353

[5] Michael Braun, Anja Mainz, Ronee Chadowitz, Bastian Pfleging, and Florian Alt. 2019. At your service: Designing voice assistant personalities to improve automotive user interfaces. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. ACM, New York, NY, USA, 1–11. http://doi.org/10.1145/3290605.3300270

[6] Barry Brown, Mathias Broth, and Erik Vinkhuyzen. 2023. The Halting problem: Video analysis of self-driving cars in traffic. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*. ACM, New York, NY, USA, 1–14. http://doi.org/10.1145/3544548.3581045

[7] Marine Capallera, Leonardo Angelini, Quentin Meteier, Omar Abou Khaled, and Elena Mugellini. 2023. Human-Vehicle Interaction to Support Driver's Situation Awareness in Automated Vehicles: A Systematic Review. *IEEE Transactions on Intelligent Vehicles* 8, 3 (March 2023), 2551–2567. http://doi.org/10.1109/TIV.2022.3200826

[8] Yilun Chen, Shah Khalid Khan, Nirajan Shiwakoti, Peter Stasinopoulos, and Kayvan Aghabayk. 2023. Analysis of Australian public acceptance of fully automated vehicles by extending technology acceptance model. *Case Stud. Transp. Policy* 14, 101072 (Dec. 2023), 101072. http://doi.org/10.1016/j.cstp.2023.101072

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

[10] Nick Gang, Srinath Sibi, Romain Michon, Brian Mok, Chris Chafe, and Wendy Ju. 2018. Don't Be Alarmed: Sonifying Autonomous Vehicle Perception to Increase Situation Awareness. In *Proceedings of the 10th International Conference on Automotive User Interfaces and Interactive Vehicular Applications (AutomotiveUI '18)*. Association for Computing Machinery, New York, NY, USA, 237–246. https://doi.org/10.1145/3239060.3265636

[11] P A Hancock, Paula A Desmond, and Gerald Matthews. 2017. Conceptualizing and Defining Fatigue. In *The Handbook of Operator Fatigue* (1st edition ed.). CRC Press, London, 63–73. http://doi.org/10.1201/9781315557366-4

[12] Christian Heath, Jon Hindmarsh, and Paul Luff. 2010. *Video in Qualitative Research*. SAGE Publications, Thousand Oaks, CA. http://doi.org/10.4135/9781526435385

[13] Shaoshuai Huang, Xuandong Zhao, Dapeng Wei, Xinheng Song, and Yuanbo Sun. 2024. Chatbot and fatigued driver: Exploring the use of LLM-based voice assistants for driving fatigue. In *Extended Abstracts of the CHI Conference on Human Factors in Computing Systems*. ACM, New York, NY, USA, 1–8. http://doi.org/10.1145/3613905.3651031

[14] Nathan Hutchins and Loyd Hook. 2017. Technology acceptance model for safety critical autonomous transportation systems. In *2017 IEEE/AIAA 36th Digital Avionics Systems Conference (DASC)*. IEEE, St. Petersburg, Florida, USA, 1–5. http://doi.org/10.1109/DASC.2017.8102010

[15] Pouya M Khorsandi, Lee Jones, Vandad Davoodnia, Timothy J Lampen, Aliya Conrad, Ali Etemad, and Sara Nabil. 2023. FabriCar: Enriching the user experience of in-car media interactions with ubiquitous vehicle interiors using E-textile sensors. In *Proceedings of the 2023 ACM Designing Interactive Systems Conference*. ACM, New York, NY, USA, 1438–1456. https://doi.org/10.1145/3563657.3595988

[16] Udo Kuckartz. 2019. Qualitative text analysis: A systematic approach. In *ICME-13 Monographs*. Springer International Publishing, Cham, 181–197. http://doi.org/10.1007/978-3-030-15636-7_8

[17] David R. Large, Gary Burnett, Vicki Antrobus, and Lee Skrypchuk. 2019. Stimulating Conversation: Engaging Drivers in Natural Language Interactions with an Autonomous Digital Driving Assistant to Counteract Passive Task-Related Fatigue. In *5th International Conference on Driver Distraction and Inattention (DDI2017)*. IFSTTAR, Paris, France, 1–17. https://nottingham-repository.worktribe.com/output/2034939Task-Related_Fatigue

[18] David R Large, Leigh Clark, Gary Burnett, Kyle Harrington, Jacob Luton, Peter Thomas, and Pete Bennett. 2019. It's small talk, jim, but not as we know it: engendering trust through human-agent conversation in an autonomous, self-driving car. In *Proceedings of the 1st International Conference on Conversational User Interfaces*. ACM, New York, NY, USA, 1–7. http://doi.org/10.1145/3342775.3342789

[19] Grégoire S Larue, Andry Rakotonirainy, and Anthony N Pettitt. 2011. Driving performance impairments due to hypovigilance on monotonous roads. *Accid. Anal. Prev.* 43, 6 (Nov. 2011), 2037–2046. http://doi.org/10.1016/j.aap.2011.05.023

[20] Joonbum Lee, Vindhya Venkatraman, John L Campbell, and Christian M Richard. 2019. Workload and attention management in automated vehicles. In *Human Performance in Automated and Autonomous Systems* (1st edition ed.). CRC Press, Boca Raton, FL : CRC Press/Taylor & Francis Group, 2019., 213–230. https://doi.org/10.1201/9780429458330

[21] Nadine Lessio and Alexis Morris. 2020. Toward design archetypes for conversational agent personality. In *2020 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*. IEEE, Toronto, ON, Canada, 3221–3228. http://doi.org/10.1109/SMC42975.2020.9283254

[22] Alexandra Loew, Christina Kurpiers, Martin Götze, Sven Nitsche, and Klaus Bengler. 2024. How to counteract driver fatigue during conditional automated driving—A systematic review. *Future Transportation* 4, 1 (March 2024), 283–298. http://doi.org/10.3390/futuretransp4010015

[23] Kirti Mahajan, David R Large, Gary Burnett, and Nagendra R Velaga. 2021. Exploring the benefits of conversing with a digital voice assistant during automated driving: A parametric duration model of takeover time. *Transp. Res. Part F Traffic Psychol. Behav.* 80 (July 2021), 104–126. https://doi.org/10.1016/j.trf.2021.03.012

[24] Jennifer F May and Carryl L Baldwin. 2009. Driver fatigue: The importance of identifying causal factors of fatigue when considering detection and countermeasure technologies. *Transp. Res. Part F Traffic Psychol. Behav.* 12, 3 (May 2009), 218–224. http://dx.doi.org/10.1016/j.trf.2008.11.005

[25] Angus McKerral, Kristen Pammer, and Cassandra Gauld. 2023. Supervising the self-driving car: Situation awareness and fatigue during highly automated driving. *Accident Analysis & Prevention* 187 (2023), 107068. https://doi.org/10.1016/j.aap.2023.107068

[26] Thomas McWilliams and Nathan Ward. 2021. Underload on the road: Measuring vigilance decrements during partially automated driving. *Front. Psychol.* 12 (April 2021), 631364. http://doi.org/10.3389/fpsyg.2021.631364

[27] Shashank Mehrotra, Meng Wang, Nicholas Wong, Jah'inaya Parker, Shannon C Roberts, Woon Kim, Alicia Romo, and William J Horrey. 2022. Human-Machine Interfaces and Vehicle Automation: A Review of the Literature. *Accid. Anal. Prev.* 109 (2022), 18–28. https://doi.org/10.1016/j.trf.2024.08.014

[28] Natasha Merat, Bobbie Seppelt, Tyron Louw, Johan Engström, John D Lee, Emma Johansson, Charles A Green, Satoshi Katazaki, Chris Monk, Makoto Itoh, Daniel McGehee, Takashi Sunda, Kiyozumi Unoura, Trent Victor, Anna Schieben, and Andreas Keinath. 2019. The "Out-of-the-Loop" concept in automated driving: proposed definition, measures and implications. *Cogn. Technol. Work* 21, 1 (Feb. 2019), 87–98. https://doi.org/10.1007/s10111-018-0525-8

[29] Alireza Mortezapour. 2025. Large language models in the service of ergonomics education: a theoretical discussion on extending the ergonomics curriculum through prompt engineering. *Theor. Issues Ergon.* 0, 0 (Oct. 2025), 1–11. http://doi.org/10.1080/1463922X.2025.2568125

[30] Frederik Naujoks, Simon Höfling, Christian Purucker, and Kathrin Zeeb. 2018. From partial and high automation to manual driving: Relationship between non-driving related tasks, drowsiness and take-over performance. *Accid. Anal. Prev.* 121 (Dec. 2018), 28–42. http://doi.org/10.1016/j.aap.2018.08.018

[31] Michael A Nees, Benji Helbein, and Anna Porter. 2016. Speech Auditory Alerts Promote Memory for Alerted Events in a Video-Simulated Self-Driving Car Ride. *Hum. Factors* 58, 3 (May 2016), 416–426. https://doi.org/10.1177/0018720816629279

[32] Lene Nielsen. 2019. *Personas - user focused design* (2 ed.). Springer, London, England. http://doi.org/10.1007/978-1-4471-7427-1

[33] Donggun Park, Yushin Lee, and Yong Min Kim. 2024. Effects of autonomous driving context and anthropomorphism of in-vehicle voice agents on intimacy, trust, and intention to use. *Int. J. Hum. Comput. Interact.* 40, 22 (Nov. 2024), 7179–7192. http://doi.org/10.1080/10447318.2023.2262271

[34] Robert Plutchik. 1982. A psychoevolutionary theory of emotions. *Soc. Sci. Inf. (Paris)* 21, 4-5 (July 1982), 529–553. http://doi.org/10.1177/053901882021004003

[35] Eileen Roesler, Lara Naendrup-Poell, Dietrich Manzey, and Linda Onnasch. 2022. Why context matters: The influence of application domain on preferred degree of anthropomorphism and gender attribution in human–robot interaction. *Int. J. Soc. Robot.* 14, 5 (July 2022), 1155–1166. http://doi.org/10.1007/s12369-021-00860-z

[36] Peter A M Ruijten, Jacques M B Terken, and Sanjeev N Chandramouli. 2018. Enhancing trust in autonomous vehicles through intelligent user interfaces that mimic human behavior. *Multimodal Technol. Interact.* 2, 4 (Sept. 2018), 62. http://doi.org/10.3390/mti2040062

[37] Shadan Sadeghian Borojeni, Susanne C J Boll, Wilko Heuten, Heinrich H Bülthoff, and Lewis Chuang. 2018. Feel the Movement: Real Motion Influences Responses to Take-over Requests in Highly Automated Vehicles. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems (CHI '18, Paper 246)*. Association for Computing Machinery, Montreal QC Canada, 1–13. https://doi.org/10.1145/3173574.3173820

[38] SAE International. 2021. *Taxonomy and Definitions for Terms Related to Driving Automation Systems for On-Road Motor Vehicles*. Technical Report. SAE International.

[39] Dyani J Saxby, Gerald Matthews, Edward M Hitchcock, Joel S Warm, Gregory J Funke, and Thomas Gantzer. 2008. Effect of active and passive fatigue on performance using a driving simulator. *Proc. Hum. Factors Ergon. Soc. Annu. Meet.* 52, 21 (Sept. 2008), 1751–1755. http://doi.org/10.1177/154193120805202113

[40] B Schilit, N Adams, and R Want. 1994. Context-aware computing applications. In *1994 First Workshop on Mobile Computing Systems and Applications*. IEEE, Santa Cruz, CA, 85–90. http://doi.org/10.1109/WMCSA.1994.16

[41] Ronald Schroeter and Fabius Steinberger. 2016. Pokémon DRIVE: towards increased situational awareness in semi-automated driving. In *Proceedings of the 28th Australian Conference on Computer-Human Interaction* (Launceston, Tasmania, Australia) *(OzCHI '16)*. Association for Computing Machinery, New York, NY, USA, 25–29. https://doi.org/10.1145/3010915.3010973

[42] Annika Stampf, Mark Colley, and Enrico Rukzio. 2022. Towards implicit interaction in highly automated vehicles - A systematic literature review. *Proc. ACM Hum. Comput. Interact.* 6, MHCI (Sept. 2022), 1–21. https://doi.org/10.1145/3546726

[43] Fabius Steinberger, Ronald Schroeter, and Christopher N Watling. 2017. From road distraction to safe driving: Evaluating the effects of boredom and gamification on driving behaviour, physiological arousal, and subjective experience. *Comput. Human Behav.* 75 (Oct. 2017), 714–726. http://doi.org/10.1016/j.chb.2017.06.019

[44] Abbas. Tashakkori and Charles. Teddlie. 2010. *Sage handbook of mixed methods in social & behavioral research* (2nd ed. ed.). SAGE Publications, Los Angeles.

[45] Tom van Veen, Juffrizal Karjanto, and Jacques Terken. 2017. Situation awareness in automated vehicles through proximal peripheral light signals. In *Proceedings of the 9th International Conference on Automotive User Interfaces and Interactive Vehicular Applications*. Association for Computing Machinery, Oldenburg Germany, 287–292. https://doi.org/10.1145/3122986.3122993

[46] Viswanath Venkatesh and Hillol Bala. 2008. Technology acceptance model 3 and a research agenda on interventions. *Decis. Sci.* 39, 2 (May 2008), 273–315. https://doi.org/10.1111/j.1540-5915.2008.00192.x

[47] Tobias Vogelpohl, Matthias Kühn, Thomas Hummel, and Mark Vollrath. 2019. Asleep at the automated wheel-Sleepiness and fatigue during highly automated

driving. *Accid. Anal. Prev.* 126 (May 2019), 70–84. http://dx.doi.org/10.1016/j.aap.2018.03.013

[48] Veronika Weinbeer, Tobias Muhr, and Klaus Bengler. 2019. Automated driving: The potential of non-driving-related tasks to manage driver drowsiness. In *Advances in Intelligent Systems and Computing*. Springer International Publishing, Cham, 179–188. http://doi.org/10.1007/978-3-319-96074-6_19

[49] Mark S Young and Neville A Stanton. 2002. Malleable Attentional Resources Theory: A New Explanation for the Effects of Mental Underload on Performance. *Hum. Factors* 44, 3 (Sept. 2002), 365–375. https://doi.org/10.1518/0018720024497709

[50] Yueteng Yu, Lewis Cockram, Jorge Pardo Gaytan, Xiaomeng Li, Andry Rakotonirainy, Jonny Kuo, Sebastien Demmel, Mike Lenné, and Ronald Schroeter. 2025. Leveraging LLM-based Conversational Agents to Combat Passive Fatigue in Conditional Automated Driving Contexts. In *Proceedings of the 17th International Conference on Automotive User Interfaces and Interactive Vehicular Applications*. Association for Computing Machinery (ACM), Brisbane, AU, 1–5. https://doi.org/10.1145/3744335.3758493

[51] Yueteng Yu, Xiaomeng Li, Sebastien Demmel, Sebastien Glaser, Jonny Kuo, Mike Lenné, and Ronald Schroeter. 2025. Designing With Motion: Exploring Vestibular Cues as a Subtle Awareness Nudge Modality in Automated Vehicles. In *17th International Conference on Automotive User Interfaces and Interactive Vehicular Applications (AutomotiveUI '25)*. Association for Computing Machinery (ACM), United States, 1–11. https://doi.org/10.1145/3744333.3747824

Table 2: Appendix: Participants' demographic table

PID	Gender	Age	Driving Exp.	Driving Frequency	AV Exp. (L1+)	CA/NCA
P01	Male	43	5+ yrs	5-7 days per week	Yes	CA
P02	Female	73	5+ yrs	At least once in the last year	No	CA
P03	Female	48	5+ yrs	3-4 days per week	No	CA
P04	Female	40	5+ yrs	5-7 days per week	No	CA
P05	Male	31	5+ yrs	5-7 days per week	Yes	CA
P06	Male	72	5+ yrs	5-7 days per week	No	NCA
P07	Female	31	2-5 yrs	At least once every month	Yes	CA
P08	Female	31	5+ yrs	At least once every month	No	CA
P12	Male	38	5+ yrs	1-2 days per week	Yes	CA
P16	Female	44	5+ yrs	3-4 days per week	No	CA
P22	Male	32	5+ yrs	1-2 days per week	No	CA
P25	Male	32	5+ yrs	3-4 days per week	Yes	CA
P29	Male	44	5+ yrs	3-4 days per week	Yes	NCA
P31	Male	34	5+ yrs	At least once in the last year	No	NCA
P36	Female	32	5+ yrs	5-7 days per week	No	NCA
P39	Female	33	5+ yrs	At least once every month	No	CA
P43	Male	77	5+ yrs	5-7 days per week	No	CA
P49	Female	81	5+ yrs	5-7 days per week	No	NCA
P50	Female	81	5+ yrs	5-7 days per week	No	CA
P62	Male	69	5+ yrs	3-4 days per week	Yes	CA
P64	Male	67	5+ yrs	5-7 days per week	No	NCA
P65	Male	85	5+ yrs	3-4 days per week	Yes	CA
P79	Male	73	5+ yrs	5-7 days per week	No	NCA
P80	Female	72	5+ yrs	1-2 days per week	No	NCA
P84	Female	82	5+ yrs	1-2 days per week	No	CA
P85	Male	74	5+ yrs	5-7 days per week	No	CA
P88	Female	68	5+ yrs	5-7 days per week	No	NCA
P89	Male	82	5+ yrs	5-7 days per week	No	CA
P94	Male	36	5+ yrs	3-4 days per week	Yes	NCA
P95	Female	67	5+ yrs	5-7 days per week	No	CA
P99	Male	71	5+ yrs	5-7 days per week	No	CA
P100	Female	33	5+ yrs	5-7 days per week	Yes	NCA
P104	Male	73	5+ yrs	3-4 days per week	Yes	CA
P106	Male	50	5+ yrs	5-7 days per week	No	NCA
P107	Male	66	5+ yrs	5-7 days per week	No	NCA
P108	Male	35	5+ yrs	5-7 days per week	Yes	NCA
P110	Male	47	5+ yrs	5-7 days per week	No	NCA
P111	Female	77	5+ yrs	3-4 days per week	No	CA
P112	Male	66	5+ yrs	5-7 days per week	No	CA
P113	Male	80	5+ yrs	5-7 days per week	No	CA