International Journal of Artificial Intelligence in Education (2025) 35:736–773
https://doi.org/10.1007/s40593-024-00407-z

ARTICLE

Transforming Driver Education: A Comparative Analysis of LLM-Augmented Training and Conventional Instruction for Autonomous Vehicle Technologies

Mohsin Murtaza[1] · Chi-Tsun Cheng[1] · Mohammad Fard[1] · John Zeleznikow[2]

Accepted: 14 April 2024 / Published online: 14 May 2024

Abstract
As modern vehicles continue to integrate increasingly sophisticated Advanced Driver Assistance Systems (ADAS) and Autonomous Vehicles (AV) functions, conventional user manuals may no longer be the most effective medium for conveying knowledge to drivers. This research analysed conventional, paper and video-based instructional methods versus a Large Language Model (LLM)-based instructional tool to educate 86 participants about the operation of specific ADAS and AV functionalities. The study sampled participants aged between 20 and over 40, with driving experience ranging from one to over six years. The first group was educated using the conventional methods. In contrast, the second group received instructions via an LLM, i.e., users learn via ChatGPT interaction. Our goal was to assess the efficiency and effectiveness of these teaching methodologies based on the reaction times participants required to activate ADAS functions and the corresponding accuracies. Our findings revealed that the group trained via ChatGPT demonstrated significantly improved learning outcomes compared to conventional training. This included shorter activation times, higher consistency, and higher accuracy across examined functions. This study further proposed a framework to effectively use ChatGPT for different training scenarios and education purposes, offering a valuable resource for leveraging Artificial Intelligence (AI) in training users to handle complex systems. The framework empowers educators to tailor ChatGPT's interactions, ensuring efficient, guided learning experiences for learners. For researchers, this study lays the foundation for exploring the role of LLM-based instructional tools in a broader range of applications.

Keywords Advanced driver assistance systems (ADAS) · Artificial intelligence (AI) · Autonomous vehicles (AV) · ChatGPT · Driver training · Large language model (LLM)

Extended author information available on the last page of the article

Introduction

The rise and rapid proliferation of Advanced Driver Assistance Systems (ADAS) and Autonomous Vehicles (AV) constitute a significant transformation in the automotive industry, possessing the potential to reshape transportation systems across the globe drastically (Fagnant & Kockelman, 2015; Litman, 2017) and offer various benefits, notably increased safety (Zahabi et al., 2020), enhanced mobility, and a substantial reduction in traffic congestion (Fagnant & Kockelman, 2015). Human factors contribute to more than 90% of accidents, according to Australia's National Road Safety Partnership Program (NRSPP) (Murtaza et al., 2023). The technology-driven shifts triggered by the advancements in ADAS, and AVs can significantly mitigate human-related accidents, which remain one of the leading causes of road casualties worldwide. The capabilities of AVs vary based on their level of automation, as categorised by the Society of Automotive Engineers (SAE), ranging from no driving automation (level 0) to full automation (level 5) (SAE International, 2021). However, the introduction of these sophisticated systems raises a fundamental question: How can human drivers be effectively trained to safely interact and cooperate with vehicles with ADAS functions or AVs (Murtaza et al., 2023)?

The seamless operation and utilisation of ADAS functions and AVs largely depend upon drivers' comprehensive understanding and ability to control these advanced systems effectively, which could be obstructed by several barriers. These include the lack of standardisation across manufacturers (Murtaza et al., 2022a) and the lack of specific training, practising platforms and opportunities (Murtaza et al., 2023). The urgency of addressing these concerns is highlighted by the growing consensus among the research community about the critical role of appropriate training in availing the full potential of AV technology (Zahabi et al., 2020; Zheng et al., 2023).

The automotive industry and academia have invested considerable resources into training to ensure trainees have the competencies to perform their tasks safely and effectively (Merriman et al., 2023). Existing training programs primarily focus on fostering confidence and imparting necessary skills in individuals, enabling them to interact with advanced systems with high proficiency and safety (Merriman et al., 2023a; Nandavar et al., 2023). The range of conventional training methods spans from paper-based, video-based instructions to demonstration-based and trial-and-error techniques.

This variance in training methods is evident across sectors. For instance, modern current vehicle driver training in the automotive industry relies on user manuals, demonstrations at dealerships, videos, information brochures, and trial and error (Boelhouwer et al., 2020; Greenwood et al., 2022; Murtaza et al., 2023; Nandavar et al., 2023; Zahabi et al., 2020), while forklift driving and situational awareness cycling also employ video-based instruction (Lehtonen et al., 2017, 2021). In the aviation industry, pilot training employs simulation-based and video-based methods (Nasir & Bargstädt, 2017; Ng, 2022; Salas et al., 1998), while the software development industry utilises both video and paper-based methods (Lloyd & Robertson, 2012; Van der Meij & Van Der Meij, 2014). Similarly, video and paper-based training is used in the medical field to instruct medical procedures and train staff, including physiotherapists (Buch et al., 2014; Ji & Butterworth, 2019). This diversity in training approaches across various industries underscores the importance of tailoring instructional methods to specific sector requirements and learning objectives, emphasising the evolving nature of educational techniques in response to technological advancements and industry-specific needs.

Understanding drivers' interaction with ADAS functions and AVs is grounded in mental models, which represent drivers' knowledge and comprehension of these advanced in-vehicle systems (Gaspar et al., 2020). These mental models can be shaped by different formal and/or informal training methods, such as trial-and-error, user manuals (Lubkowski et al., 2021), dealership demonstrations (Nandavar et al., 2023), and video-based training (Zahabi et al., 2020). An accurate mental model allows drivers to utilise ADAS effectively and safely, while an inaccurate model can lead to misuse, over-reliance, and potentially dangerous situations (Nandavar et al., 2023). Hence, choosing effective training methods is crucial in developing accurate mental models, enhancing drivers' understanding of ADAS functions, and ensuring safer interaction with these systems (Murtaza et al., 2023).

Currently, there is a lack of a formal training platform specifically designed for drivers of modern vehicles equipped with ADAS functions. According to (Kay, 2023), the recent launch of ChatGPT has been a historic breakthrough for the application of Artificial Intelligence (AI) within the realm of education. Several studies (Abdelghani et al., 2023; Al Kahf et al., 2023; Aleven et al., 2023; Biswas, 2023; Colabianchi et al., 2022; Du et al., 2023; Dwivedi et al., 2023; Fauzi et al., 2023; Firat, 2023; Nick, 2023; Sallam, 2023; Smutny & Schreiberova, 2020; Susnjak, 2023; Yuan, 2023; Zhou et al., 2021) have evidenced the potential of Large Language Models (LLMs), such as ChatGPT, in enhancing learning outcomes, learner motivation, engagement, customised support and output presentation, consistency and scalability in training across a wide range of industries including automotive, transportation, aviation, maritime, medical, education, information system, and construction. However, to our knowledge, no comprehensive study discusses how an LLM-augmented approach can be employed as an effective training tool for ADAS functions or AVs.

Consequently, this study seeks to investigate this unexplored area by offering a systematic framework for using ChatGPT's LLM as a training tool for drivers to use ADAS functions and AVs in the future. By leveraging our LLM-augmented approach's cognitive and dialogic capabilities, we aim to enhance users' understanding of ADAS, thereby improving their interactions with these systems and ensuring safer and more efficient driving experiences. The efficacy of the ChatGPT-based training approach was evaluated in an empirical study and compared with conventional training methods for operating ADAS functions. Participants' performance, as measured by their accuracies and reaction times in interacting with ADAS functions, served as a primary metric for assessing the effectiveness of different training methods. By investigating the cognitive underpinnings of learning and understanding ADAS functions, this research attempts to reveal how different training methods can influence the formation of mental models, thereby impacting the utilisation of ADAS functions.

In this study, we utilised ChatGPT's capabilities to develop an interactive training program for teaching drivers how to use ADAS functions and AV capabilities. ChatGPT was asked to digest the contents of simulated ADAS and AV manuals and convey this information through a conversational, interactive medium that adapts to users' responses. The training content was developed by inputting manual descriptions into ChatGPT. This information was transformed into clear, conversational language suitable for learners even with little or no relevant technical backgrounds. By doing so, ChatGPT provided personalised guidance, clarified user queries, and verified understanding through targeted follow-up questions. This methodology was designed to impart knowledge and engage users in a two-way dialogue, fostering a deeper understanding of ADAS functionalities and enhancing the learning experience. To ensure the consistency and accuracy of the information provided, in the experiment, ChatGPT was restricted to using only the data from the manuals, avoiding any external knowledge sources. The technical implementation of this approach utilised the foundation of instructional design and the science of learning to enhance the effectiveness of the ChatGPT-facilitated training.

The findings of this study hold implications for both the automotive industry and educators, suggesting practical strategies for ADAS training that can be incorporated into driver education programs, dealership demonstrations, and user manuals. As the automotive landscape shifts towards increased automation, it becomes imperative to re-evaluate our training methodologies. This study offers a novel perspective by integrating LLMs into the training process, providing a promising avenue for enhancing drivers' comprehension of ADAS functions and ensuring safer road experiences in an era of automation.

The remainder of this paper is organised as follows: Sect. 2 presents the methodology, experimental setup, and driving environment. Section 3 describes the recruitment process and the composition of the participants. In the same section, we also discussed the framework for effectively using the ChatGPT prompt for training purposes. Section 4 compares the two different training methods, conventional vs. ChatGPT, by analysing their corresponding participants' accuracy and response times when interacting with ADAS functions. Section 5 compares various training methods and the significance of ChatGPT-based training and its applications in other industries. Based on the findings and analyses, recommendations for stakeholders and concluding remarks are provided in Sect. 6.

Methodology

Experiment Setup

The experiment was conducted using a York driving simulator, interfaced with a Logitech G27 racing wheel system comprising pedals and a shifter module. An array of ADAS functions (detailed in Table 1) were assigned to specific buttons on the steering wheel and the shifter module to mimic an authentic driving environment.

The experimental methodology encompassed manoeuvres of an AV in a three-dimensional virtual setting. This vehicle was either autonomous or controlled by the participant, with ADAS functions in either an activated or deactivated state. The simulated AV was equipped with an automatic transmission feature, enabling the panel on the shifter module to be dedicated entirely to the activation and deactivation of

ADAS functions. Consequently, the shifter lever was disabled for the duration of the experiment.

To enhance the ecological validity of the experiment, thorough attention to detail was exerted in the design process to emulate a highly authentic and immersive driving experience. This setup is visually represented in Figs. 1 and 2. Figure 1 identifies components "a" to "c" as the steering wheel, pedals, and driving seat, respectively. Figure 2 displays components "d" to "h", signifying distinct aspects of the setup and their corresponding ADAS functions.

In Fig. 2, component "d" illustrates the location of the Lane Keeping Assist (LKA) button on the shifter module. Component "e" shows the positioning of the Autopilot-On (AP-On) function on the steering wheel. The Adaptive Cruise Control (ACC) function, represented by component "f", is positioned on the shifter module. Component "g" indicates the location of the Autopilot-Off (AP-Off) function on the steering wheel, and component "h" highlights the Collision Avoidance (CA) function, also located on the steering wheel. This exhaustive labelling and description provide a comprehensive overview of the experimental setup, contributing to the experiment's reproducibility.

Driving Scenario and Environment

The driving simulator's interface is shown in Fig. 4, where the red box stands for the front of the vehicle, i.e., the bonnet. The dashboard, which includes a speedometer

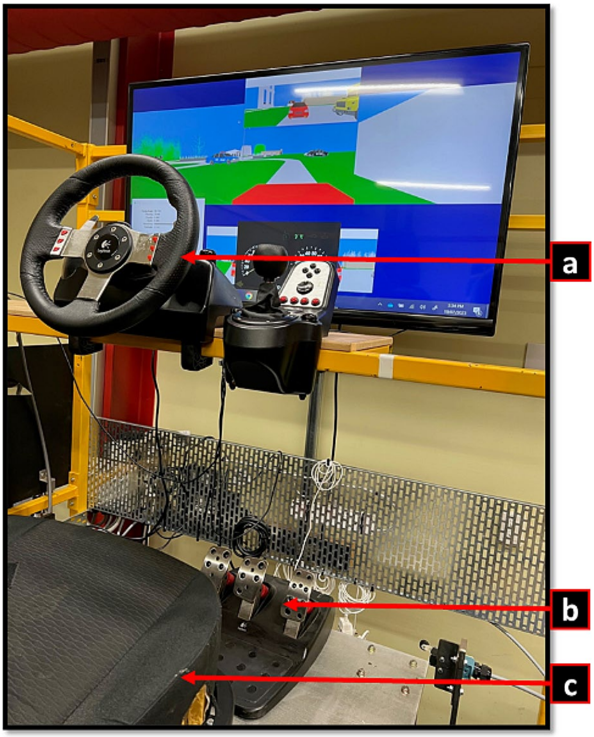


Fig. 1 Experiment setup [**a**: steering wheel, **b**: pedals, **c**: driving seat]

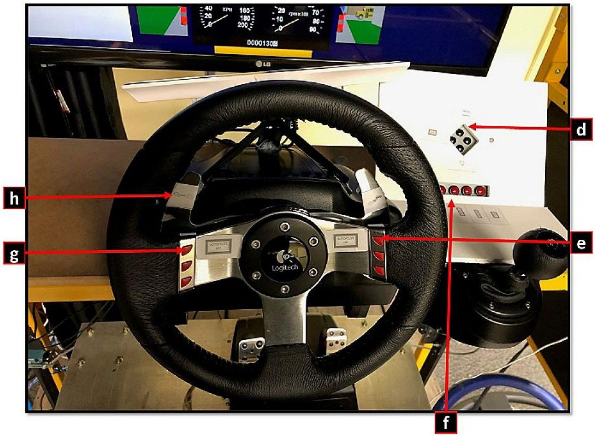


Fig. 2 ADAS functions' locations [**d**: LKA, **e**: AP-On, **f**: ACC, **g**: AP-Off, **h**: CA]

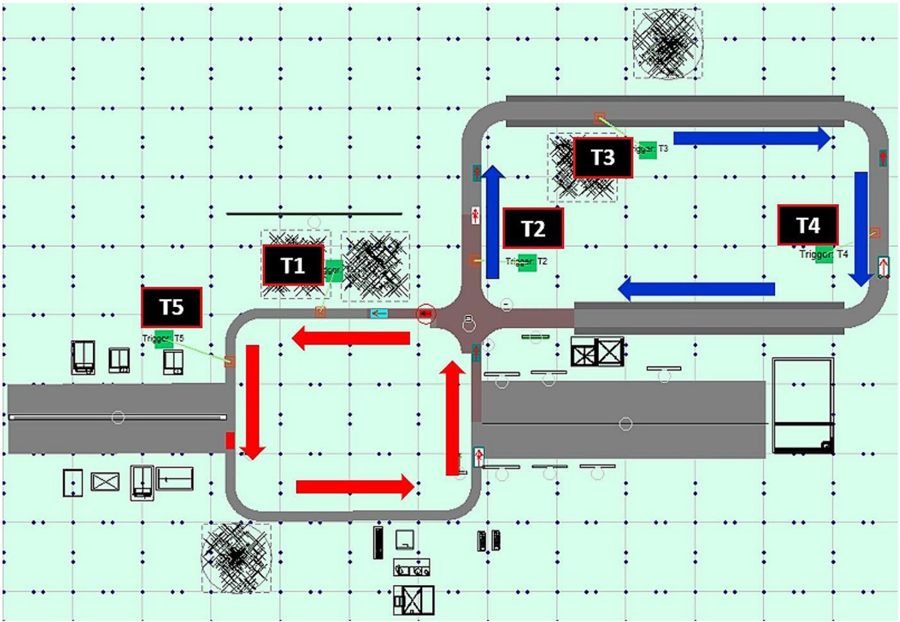


Fig. 3 An overview of the simulated driving scenario [red and blue arrows indicate the low-speed and high-speed zones, respectively]

and indicators for the ADAS functions, is displayed underneath that. These indicators light up in green to show when the functions are ON, as illustrated in Fig. 4, indicating, for example, that the LKA function is active. The 3D simulated environment incorporates several key features that one can find on a real road, such as different traffic densities, streetlights, road signs, pedestrians, speed zones, buildings, and vegetation. These features help participants better understand their environment within the simulation (Kolekar et al., 2020).

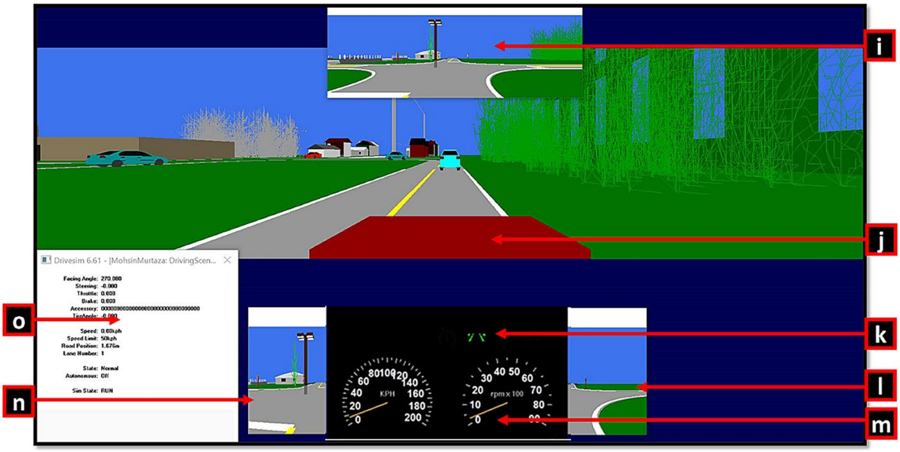


Fig. 4 A snapshot of the simulator interface [**i**: Rear mirror view, **j**: front of the simulated AV, **k**: Shows LKA function is ON, **l**: right side mirror, **m**: speedometer, **n**: left side mirror view, **o**: AV current parameters]

Table 1 Instruction for each trigger

Trigger no	Instruction
T1	Turn on the auto-pilot function
T2	Turn off the auto-pilot function
T3	Turn on the lane-keeping assist function
T4	Turn on the collision avoidance function
T5	Turn on the adaptive cruise control function

Figure 3 shows a bird's-eye view of the driving route, with a red circle showing where the vehicle starts. The route's low-speed (50 km/h) and high-speed (100 km/h) zones are indicated with red and blue arrows, respectively. These zones mimic city and highway driving changes, providing participants with different driving experiences. During each test, participants were told to start from the low-speed zone, cross an intersection, enter the high-speed zone, cross the intersection again, and then return to the low-speed zone. Several triggers (T1 to T5) along the route start different events as the vehicle passes, as shown in Fig. 3. Table 1 shows the instructions linked to these triggers. When the vehicle passes a trigger, an audio instruction from Table 1 is played, asking the participant to act accordingly. Out of the five actions, only the ACC requires two buttons to be pressed sequentially, and the others can be activated with one button. In this work, we analyse data on the steering angle, the utilisation of acceleration and braking, the times when events were triggered, and drivers' response time to these events. We measured how accurately (as a percentage) participants used the correct ADAS functions and how long they took to respond (in seconds).

Table 2 Participants division w.r.t age group

Training type	Number of participants	Age group(20–30 years)	Age group(30–40 years)	Age group (Above 40 years)
Conventional learning group	46	20	14	12
ChatGPT-based learning group	40	18	12	10

Table 3 Participants division w.r.t driving experience

Training type	Number of participants	Driving experience Novice driver (1–3 years)	Driving experience Intermediate driver (4–6 years)	Driving experience (Experienced driver (above 6 years)
Conventional learning group	46	14	13	19
ChatGPT-based learning group	40	13	11	16

Table 4 Participant's division w.r.t ADAS functions frequency of use

Training type	Number of participants	Novice/Occasional user	Intermittent user	Regular user
Conventional learning group	46	15	17	14
ChatGPT-based learning group	40	13	15	12

Recruitment of Participants

The sample comprised various ages and genders to maintain heterogeneity. The participant recruitment was conducted via a hybrid approach, utilising both digital and conventional methods. In this study, we recruited a diverse sample of participants, including RMIT University students, staff, and members of the general public interested in AV research. To recruit participants, we leveraged the university's online platform, including webpages, emails, and online forms. This effort was complemented by paper-based advertisements, i.e., posters and leaflets, distributed throughout RMIT University's Bundoora campus.

Data collection was extended over six months to ensure a sufficiently large and representative dataset for subsequent analyses. Our final sample group consisted of 86 adult participants aged 20 to above 40 years old, as detailed in Table 2. All participants had driving experience ranging from novice (one to three years), intermediate (four to six years), to experienced (over six years), as outlined in Table 3. This division of driving experience categorisation is consistent with established road safety and insurance benchmarks for defining driver expertise (Commonwealth of Massachusetts, 2023; Robbins & Chapman, 2019). Additionally, participants were categorised based on their frequency of ADAS function usage, and this information is presented in Table 4. In this study, we only collected the above information in ranges instead of their absolute values due to privacy concerns, which were essential for our ethics approval application. In the experiment, participants were divided into two distinct learning groups. They were randomly allocated to each group to enhance the validity of the comparison between training methods. Group 1, the conventional learning group, consisted of 46 participants trained through video presentations and user manuals. Group 2 comprised 40 participants who received training through a ChatGPT-based learning platform. Both groups had an almost balanced gender distribution, with the first group having 26 males and 20 females and the second group comprising 23 males and 17 females. A prerequisite for all participants was to hold a valid driving license.

To prevent potential bias, we carefully chose student participants from a wide array of academic disciplines, thereby ensuring a diverse representation of educational backgrounds. Additionally, we extended invitations to the general public, attracting participants from various backgrounds and enhancing our dataset's diversity. Before initiating the experiment, all participants were comprehensively briefed about the study's objectives and procedures. Informed consent was obtained from all participants, ensuring their voluntary and informed participation.

Given the categorical nature of our data, we conducted Pearson's Chi-Square tests for our analysis. This test is suitable for comparing frequencies across categorical variables (Field, 2013). The Mann-Whitney U test was not considered appropriate because it is intended for continuous or ordinal data with individual scores, which does not match our dataset's characteristics (Field, 2013; MacFarland et al., 2016).

We conducted statistical analyses to assess the comparability of participant characteristics between the two learning groups. The results of Pearson Chi-Square tests revealed no statistically significant association between the training group type and gender ($p=0.927$), age distribution ($p=0.989$), driving experience levels ($p=0.979$), and ADAS use frequency ($p=0.998$). All p-values were above the standard significance threshold of $p<0.05$. These findings indicate that the observed differences between the conventional learning group and the ChatGPT-based learning group were not statistically significant, suggesting a balanced distribution of participants across key demographic and experiential variables. This equivalence between groups supports the validity of comparisons made in subsequent analyses of training effectiveness.

The experimental protocol was subjected to a thorough review process and was approved by the RMIT University Human Research Ethics Committee (Approval Number: EC 25,022). This ensured adherence to the ethical standards and guidelines incumbent on academic research.

Participant's Registration and Briefing Session

Upon arrival at the experiment site, all participants were guided through a standardised registration process. They were provided with a comprehensive explanation of the experiment's setup, objectives, and potential implications. Each participant was presented with a "Participant Information and Consent Form," which had received prior approval from the RMIT University Human Research Ethics Committee. After going through the details mentioned in the form, participants gave informed consent to participate in the study by signing the form.

Subsequently, all participants were introduced to the project through a concise video. This video aimed to set the context for the experiment rather than serving as part of the training process.

Group 1 Training: Conventional Method

Following the introductory video, group 1 participants proceeded with their training session. This session commenced with an informative video lasting approximately ten minutes, which demonstrated the project's objectives, hardware setup, virtual driving environment, route, and, notably, three of the five ADAS functions.

The video's decision to focus on only three ADAS functions was informed by the need to simulate the experience of purchasing a new vehicle, where car dealers typically showcase a limited number of key functions due to time constraints and legal limitations (Murtaza et al., 2022b). Meticulous design underscored this methodology to ensure an accurate explanation of the experiment setup and a professional demonstration of the ADAS functions. To maintain consistency across the study and ensure that all participants receive the same level of information, video recordings were utilised as the medium for training and demonstration. This approach aligns with the findings of (Ebnali et al., 2019; Zahabi et al., 2020), who suggest video-based training as one of the highly effective methods.

After the video, participants were presented with a comprehensive user manual in printed form that was structured to emulate the look and feel of an actual vehicle owner's manual while being tailored for the experimental setup. The length of this user manual was around 1100 words. This manual consisted of three sections. The first section overviews the driving simulator's interface, layout, and operations. The second section offered detailed descriptions of all five ADAS functions' functionality and activation/deactivation processes, extending beyond the three discussed in the video. The final section outlined the boundaries and limitations of each ADAS function. Participants were given 15 min to read and understand the manual, emphasising the remaining two ADAS functions not covered in the video. Time allocation to read the user manual is consistent with that mentioned in (Merriman, Revell et al., 2023). The 15-minute time allocation for reading the manual was carefully chosen to maintain uniform experimental conditions across the conventional and ChatGPT groups, accounting for consistency and logistical constraints within the study design. By standardising the duration, we aimed to minimise variability in exposure to the instructional material, thus providing a fair baseline for comparing training effectiveness across demographically diverse participants. This approach ensures that any observed differences in the outcomes can be more confidently attributed to the training methods rather than differences in the time spent with the manual. According to (Brysbaert, 2019), a non-native English speaker reads an average of 139 words per minute; hence, our manual was designed to be comfortably read in around 8 min, providing ample time for participants to review sections as needed, mainly the two ADAS functions not covered in the instructional video. After training, all participants were invited to drive the AV in a simulated environment.

Group 2 Training: ChatGPT-Based Interactive User Manual

Group 2 participants underwent a distinctive training approach. Instead of the conventional method, they were introduced to a ChatGPT-based interactive user manual, with training conducted directly on a simulator screen. While the text content of the user manual remained the same for both groups, the delivery method differed significantly. This innovative approach, utilising the ChatGPT-based manual, aimed to create a more engaging and efficient learning environment within a 15-minute training session.

The ChatGPT prompt has been specifically modified to instruct participants on using numerous ADAS functions of the simulated AV, a process comprehensively discussed in Sect. 3.3.1. It is imperative to note that there is no imposed limit on the number or length of questions that participants can put forward to ChatGPT. This open-ended approach is intended to provide users with a more expansive and flexible learning environment. Participants are thus given the liberty to ask any number of questions, regardless of their complexity or scope, within 15 min. The rationale behind this procedure is to ensure that participants have ample opportunity to explore and understand the AV's operations deeply and comprehensively within a reasonable time frame.

According to (Dwivedi et al., 2023), it has been noted that instructions delivered via ChatGPT can occasionally be vague or unrelated to the context if not properly structured. This observation aligns with the findings of (Chandra et al., 2022), who asserted that poorly constructed queries or instructions may elicit incorrect or irrelevant responses from the LLMs. Consequently, to ensure that ChatGPT provides accurate instructions tailored to a specific scenario, it becomes imperative to furnish the system with ample context and detailed information.

In this study, we proposed a comprehensive set of guidelines to optimise communication with ChatGPT. These guidelines describe the most effective strategies for structuring queries and instructions to obtain the expected and most beneficial responses from the LLM-augmented approach. By adhering to these guidelines, users can maximise the potential of ChatGPT and facilitate more accurate and contextually relevant outputs, thereby enhancing the overall effectiveness of the interactive training sessions. It is important to highlight that participants were not required to create prompts to interact with ChatGPT. This aspect was managed by providing predesigned prompts designed by the Author, as detailed in Sect. 3.3.1 of our manuscript.

Framework for Preparing a ChatGPT Prompt for Training

The development and application of instructional content through ChatGPT requires a detailed understanding of the context and background information relevant to the training scenario. This paper introduces a comprehensive framework aimed at optimising the creation and delivery of ChatGPT prompts for a variety of training domains. Although our primary focus is on ADAS and AV driver training, our framework's principles are broadly applicable, extending to diverse industries beyond AV training. The framework is structured around a set of guiding principles that enhance the instructional effectiveness of ChatGPT, ensuring that the content is both relevant and accessible to learners from different backgrounds.

Our framework's approach is designed with a dual focus, incorporating general LLM prompting principles, which apply universally across multiple contexts, and application-specific (pedagogical) principles tailored to particular training scenarios such as those involving ADAS and AV. The general principles aim to guide the creation of effective, clear, focused, and contextually appropriate prompts. The application-specific principles enable the customisation of instructional content to meet the unique needs of learners. This dual approach ensures that ChatGPT delivers informative, engaging, personalised, and directly relevant training to the specific learning objectives.

In applying these principles to AV driver training, we equipped ChatGPT with a detailed user manual for an AV, instructing it to use this information to conduct an interactive, conversation-based training session. This method focuses on personalised learning experiences and ensures that instruction is based solely on the provided materials, thereby avoiding external information that could lead to inconsistencies or inaccuracies. The framework offers a versatile methodology for developing instructional content for ChatGPT by articulating a clear distinction between general and application-specific principles. This approach facilitates the creation of training programs that are effective, engaging, and tailored to the diverse needs and backgrounds of learners. To optimise the instructional effectiveness of ChatGPT in training scenarios, such as ADAS and AV driver training, we introduce a framework distinguishing between general LLM prompting principles and application-specific (pedagogical) principles. This distinction clarifies the underlying rationale of each principle and its relevance to specific applications, including ADAS and AV training. To study the practical application and effectiveness of our framework, we present a detailed demonstration of its implementation in training drivers for AV. This illustrates the adaptability and precision of our guiding principles and also showcases the tangible benefits of our approach. This practical demonstration serves as a robust proof of concept, reinforcing the applicability of our framework across various training domains and specifically highlighting its efficacy in preparing drivers for the complexities of navigating AV.

General LLM Prompting Principles

These foundational principles are applicable across a wide range of applications and are crucial for creating effective ChatGPT prompts:

i. **Goal and scene setting**: Begin the conversation with ChatGPT by clearly defining the goals, approaches, and conditions. The rationale behind setting the scene and defining goals is to provide a clear context that guides the LLM to respond appropriately to the user's needs within the specified scenario. When the LLM is aware of the scene, it can customise its language, tone, and content to match the specific requirements of the task, ensuring that the training is relevant and effective. A well-defined prompt significantly reduces the likelihood of misunderstandings, which are critical to avoid when the LLM is the primary source of instruction. The design rational is consistent with (Chandra et al., 2022; Dwivedi et al., 2023).

Below is a practical example that demonstrates how to apply our framework's principles, explicitly focusing on the initial goal and scene setting with ChatGPT.

> Hello ChatGPT, in my next message, I will provide you with the contents of a user manual for an autonomous vehicle (AV). This manual provides information on how to operate an AV in a simulated environment. Instead of asking the participants to read the manual by themselves, you will digest the contents and then use an interactive, conversation-based approach to teach the participants how to drive the AV and operate various ADAS functions [Approaches]
> Your goal is to ensure that participants feel confident and capable of using the ADAS functions and operating the AV safely and efficiently after they have received this training from you [Goals]
> Please ensure that the guide and instructions that you provide to the participants are solely based on the information provided within the given manual and do not draw from general knowledge or external resources [Conditions]

Following this, the contents of the user manual were pasted to ChatGPT via the prompt.

In this study, we provided ChatGPT with all the contents in the user manual, which include clear definitions of the operation procedures and limitations of each ADAS function. We instructed ChatGPT not to pull information from external resources to maintain clarity and consistency. This approach ensures all participants receive the same level of detail about each function, avoiding the situation where ChatGPT might generate false information, thus reducing potential confusion.

ii. **Providing structured instruction to ChatGPT**: Consistency is crucial in helping ChatGPT understand and respond appropriately. This could apply to various scenarios, like instructing different operations in a factory or explaining other administrative procedures in an office setting. Framing information consistently can significantly improve the clarity of the response and facilitate the learning process. To ensure this, we have established a standard structure for presenting instructions to ChatGPT. For every ADAS function, we begin with its function name, then provide a brief description, specify its physical location on the experiment setup, verbally describe its symbol, and finally, with its activation and deactivation procedures. This consistent approach facilitates more effective prompt responses. Since our user manual has already been written in a systemic structure, no special instructions are needed in our experiment.
iii. **Reiterate objectives and conditions to ChatGPT regularly**: It is vital to remind ChatGPT regularly each time a session is initiated or resumed to adhere strictly to the instructions provided in the manual or training materials. It could be critical because of the conversational memory and the maximum token limitations in LLM-based tools. This practice could make ChatGPT less likely to divert or start utilising information from external or generic sources and thus

ensure context retention and adherence to specified guidelines throughout the interaction. Regular reinforcement of these guidelines helps maintain the focus and accuracy of the ChatGPT responses. Such reiteration can compensate for the absence of a sustained conversational memory (Open, 2023) in AI systems such as ChatGPT, facilitating more accurate and contextually relevant responses. However, our study did not encounter this issue, as each participant received training through separate prompts. Furthermore, the user manual was concise and short (approximately 1100 words); therefore, we did not reach ChatGPT's maximum token capacity.

Tokenisation and chunking, specifically in the context of Open-AI's GPT-3 and GPT-4, refers to breaking down the input and output into smaller pieces known as tokens. A token could be as small as one character or a word (for example, the letter 'a' could be a token, and the word 'apple' could also be a token). When there's a long dialogue with many back-and-forth conversations, ChatGPT can hit its maximum token limit (e.g., GPT maximum is 4097 tokens) (Raf, 2023). When this limit is reached, some of the earlier parts of the conversation may be cut off, resulting in the model losing the context or instructions provided at the beginning. To address this issue, important instructions should be reiterated regularly to ensure the model continues following them. This is especially crucial when the instructions involve adhering strictly to a particular set of guidelines or utilising specific terms. For instance, in this study, they are the names and symbols of the ADAS functions.

Application-Specific (Pedagogical) Principles

iv. **Personalised responses**: Customise your instructions and questions to match your task's specific requirements and your audience's characteristics. Personalised responses ensure that the instruction is comprehensible to participants from diverse backgrounds, thereby enhancing their understanding and application of the information provided. For instance, considering our participants are from diverse backgrounds, we guided ChatGPT to keep its instructions simple and clear. This makes it easier for participants to understand and follow. However, if the participating groups have relevant professional knowledge, using well-known technical terminologies and jargon can ease understanding and make the conversation more engaging. Highlighting or bolding key points may help draw participants' attention to crucial information. For example, in our experiment, we instructed ChatGPT: "*Your audience is the general public; please use simple words and highlight keywords in your response to participants' inquiries*."

v. **Proactive knowledge checking on learners**: One can train ChatGPT to respond to a query with the corresponding answer plus a relevant follow-up question to check a learner's understanding. Using follow-up questions is one of the best methods to reinforce learning through retrieval practice (Agarwal & Bain, 2019). This can be achieved by instructing ChatGPT to pose a follow-up question to the participant after its response. For instance, in this study, we instructed ChatGPT that if a participant asks how to activate any ADAS function, it should display the steps and then ask a follow-up question. The following instruction was given to ChatGPT: "*After you respond to the participant's question, please ask a follow-up question similar to the one which is mentioned below". "Do you understand how to use this function, or would you like me to explain it to you again"?*

vi. **Phrasing output format**: One can instruct ChatGPT to present its responses in various formats, such as bullet points, paragraphs, or tables. Utilising bullet points is particularly effective for presenting sequential instructions in a clear and easy-to-follow manner. For instance, in this study, we guided ChatGPT to rephrase its output by instructing it to; "*Please use bullet points to list the steps for activating or deactivating any function".*

Assessment of ChatGPT's Response Accuracy and Sample Interactions

In our experimental setup, we designed a unique user manual, approximately 1,100 words in length, detailing the location and function of each ADAS feature. Three functions were located on the steering wheel and two on the shifter module. This specificity ensured that ChatGPT relied solely on the information provided, which helped maintain the accuracy of its responses. Throughout the experiment, we did not observe any false responses generated by ChatGPT.

However, ChatGPT does have limitations, especially after pauses in interaction or when initiating a new session due to the sustained conversational memory issue in ChatGPT (Open, 2023), as well as when reaching the maximum token limit (Raf, 2023), which may result in irrelevant or incorrect information. These limitations are further compounded when requesting real-time or very recent information (Du et al., 2023). We did not observe any false or incorrect responses generated by ChatGPT during our experiment, confirming its accuracy within our specific parameters. The study's limitation lies in the brief nature of our user manual (which was around 1100 words in length), which does not adequately represent the complexities of processing larger documents. Such complexities could challenge ChatGPT's token capacity, potentially increasing the risk of generating inaccurate or invented information.

Furthermore, for more robust analysis, it would be ideal to implement systematic procedures to identify false, biased, or incomplete information automatically. For future work, it is also important to investigate how LLMs manage more lengthy documents and whether this affects the tendency to invent information. Future research should also be conducted to validate ChatGPT's performance in terms of accuracy and reliability. This aspect has also been highlighted by (Thirunavukarasu et al., 2023) as a key area for future research.

Participants learned about all the ADAS and AV functions available in the simulated AV through interactions with ChatGPT. To provide a comprehensive overview of ChatGPT's performance, we present five examples of questions asked by the learners, distributed across three types of inquiries. These include two questions relating to functions available in our AV simulator, two queries about non-existing functions, and one more complex, conditional inquiry illustrated through a specific scenario-based question. The rationale was to assess whether ChatGPT can provide valid answers across various scenarios without inventing information. In all cases,

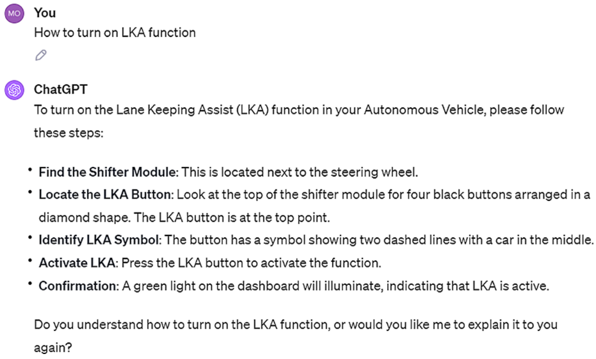


Fig. 5 Screenshot of a learner-1 inquiring about an existing function in the simulator

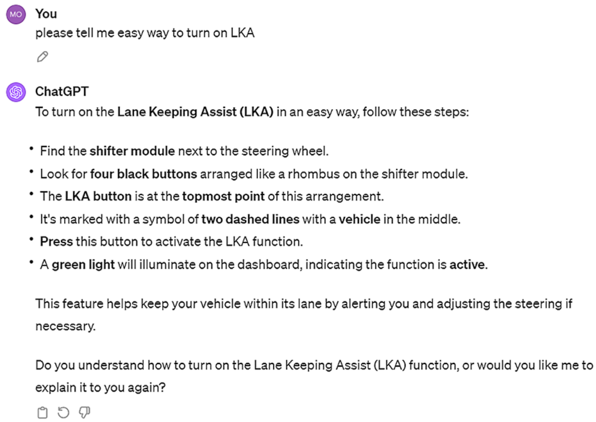


Fig. 6 Screenshot of a learner-2 inquiring about an existing function in the simulator

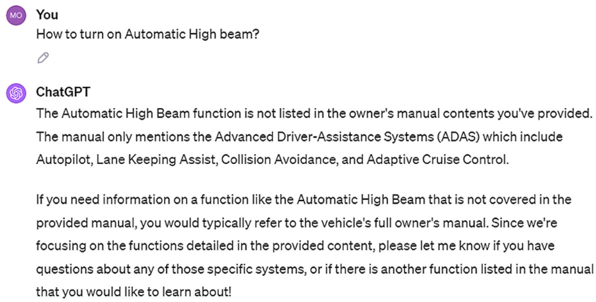
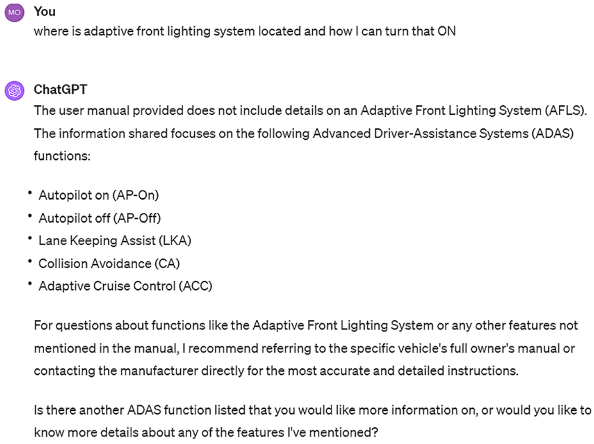


Fig. 7 Screenshot of a learner-1 inquiring about a non-existing function in the simulator

the responses from ChatGPT were accurate, demonstrating the model's capability to handle direct inquiries (the functions available in simulated AV as shown in Figs. 5 and 6) and hypothetical scenarios (functions not available in AV as demonstrated in Figs. 7 and 8), and complex, conditional inquiries (as illustrated in the new Fig. 9) effectively.

Additionally, we observed that, on average, participants asked seven questions while interacting with ChatGPT. The responses from ChatGPT, averaging approximately 124 words per question, were concise yet comprehensive, making it easier for

Fig. 8 Screenshot of a learner-2 inquiring about a non-existing function in the simulator


You
where is adaptive front lighting system located and how I can turn that ON

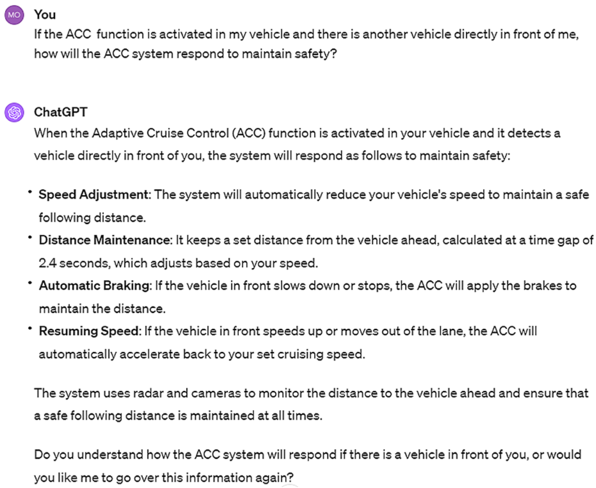
ChatGPT
The user manual provided does not include details on an Adaptive Front Lighting System (AFLS). The information shared focuses on the following Advanced Driver-Assistance Systems (ADAS) functions:

- Autopilot on (AP-On)
- Autopilot off (AP-Off)
- Lane Keeping Assist (LKA)
- Collision Avoidance (CA)
- Adaptive Cruise Control (ACC)

For questions about functions like the Adaptive Front Lighting System or any other features not mentioned in the manual, I recommend referring to the specific vehicle's full owner's manual or contacting the manufacturer directly for the most accurate and detailed instructions.

Is there another ADAS function listed that you would like more information on, or would you like to know more details about any of the features I've mentioned?

Fig. 9 Screenshot of a learner inquiring about a complex, conditional inquiry of an existing function in the simulator


You
If the ACC function is activated in my vehicle and there is another vehicle directly in front of me, how will the ACC system respond to maintain safety?

ChatGPT
When the Adaptive Cruise Control (ACC) function is activated in your vehicle and it detects a vehicle directly in front of you, the system will respond as follows to maintain safety:

- **Speed Adjustment**: The system will automatically reduce your vehicle's speed to maintain a safe following distance.
- **Distance Maintenance**: It keeps a set distance from the vehicle ahead, calculated at a time gap of 2.4 seconds, which adjusts based on your speed.
- **Automatic Braking**: If the vehicle in front slows down or stops, the ACC will apply the brakes to maintain the distance.
- **Resuming Speed**: If the vehicle in front speeds up or moves out of the lane, the ACC will automatically accelerate back to your set cruising speed.

The system uses radar and cameras to monitor the distance to the vehicle ahead and ensure that a safe following distance is maintained at all times.

Do you understand how the ACC system will respond if there is a vehicle in front of you, or would you like me to go over this information again?

participants to understand and engage with the content. Furthermore, we analysed the total length of participant's communication with ChatGPT, including the provided AV user manual, instructions and questions asked by the participants, which amounted to approximately 2520 words. This quantitative analysis is instrumental in highlighting the level of engagement, depth of user interactions and the efficacy of ChatGPT's responses.

Preparing the Learners

The effectiveness of using ChatGPT as a teaching aid in education and training depends on educators or trainers providing clear guidance to their learners on how to interact with and learn from the system. In this work, our training approach is based on a well-structured preparation process, which was delivered through a short introduction video to ensure consistency in the instruction mode. The following steps outline our methodology for preparing the video:

Defining Learning Objectives In the beginning section, we set clear learning goals, emphasising the participants' comprehension of ADAS functions. This step provides a clear target for the training and ensures learners know what they are expected to learn.

Introducing ADAS Context and Vocabulary The next section in the instructional video presents the experimental context and introduces key ADAS features within a simulated AV environment to engage the participants. Simultaneously, it provides essential ADAS-related technical terms such as Autopilot, ACC, LKA, and CA, equipping learners with the language needed for accurate discussions with ChatGPT. This ensures that the discussions are technically relevant.

An Interaction Kickstarting Example The final section of the video guides learners on initiating interaction with ChatGPT. It suggests starting the conversation with specific queries like, "*I am a new user, please tell me what functions are available in the AV"?* to ensure that learners engage in informed discussions about the functions and their operational processes.

Results and Analysis

In our study, we aimed to investigate the significance and impact of different training approaches for end-users of ADAS and autonomous driving functionalities within an AV. Participants interacted with the AV system within a simulated environment, using either an innovative training method facilitated by ChatGPT or a conventional method comprising video demonstrations and user manual instructions. A critical element of ChatGPT-based training was developing a set of principles designed to optimise the effectiveness of ChatGPT in generating prompts and yielding better learning outcomes. Performance metrics included participants' accuracies, represented by the percentage of correct actions executed, and response latencies, calculated as the time difference between when the audio instruction was issued and when the correct ADAS function was activated or deactivated.

Accuracy Analysis

The results of this study are derived from the comparative analysis of two groups: a conventional learning group of 46 participants and a ChatGPT group of 40 participants, a total of 86 participants. Participants were trained on five ADAS functions: AP-On, AP-Off, LKA, CA, and ACC, using their respective training methods.

According to the data presented in Table 5, those trained with ChatGPT demonstrated a comprehensive understanding of all five functions, achieving accuracies ranging from 80 to 100%. It was observed that both ChatGPT and video-based learn-

Table 5 Participants' average accuracies after being trained using different methods

Group-division w.*r*.t training method	No. of participants in each group	AP-on response accuracy in %	AP-off response accuracy in %	LKA response accuracy in %	CA response accuracy in %	ACC response accuracy in %
ChatGPT-based learning Group	40	100	100	92	95	80
Conventional learning group	46	100	100	71	73	58

ers reached 100% accuracy in identifying and using the AP-On and AP-Off functions, irrespective of training methods and their driving experience levels. The prominent placement of these AP functions at the top of the steering wheel likely contributed to the uniform high accuracy rates.

The physical placement of each ADAS function is also an important factor in understanding the study's outcomes. LKA and ACC functions were situated on the shifter module, whereas the CA system's control was accessible on the side of the steering wheel. Compared to the AP-On and AP-Off functions at the top of the steering wheel, the less intuitive placement of the LKA, CA, and ACC functions likely contributed to the varied accuracy rates across these functions. The ChatGPT group outperformed the video-based group for the LKA function by achieving a 92% accuracy rate compared to the latter's 71%. It is worth noting that the video-based group did not have data for CA and ACC, mimicking a real-world scenario where not all ADAS functions are demonstrated by the sales agent. The ChatGPT group demonstrated a 95% accuracy rate for the CA function, compared to the 73% observed in the user manual group. Likewise, ACC function accuracy was 80% for ChatGPT learners, compared to 58% for those who used the user manual.

To thoroughly evaluate the effectiveness of the training methods, a statistical analysis was conducted using the Pearson Chi-square method. This analysis revealed significant differences in learning outcomes between the groups for several ADAS functions. Specifically, the p-values obtained are 0.014 for the LKA function, 0.008 for the ACC function, and 0.034 for the CA function. Given that these p-values are below the threshold of 0.05, it can be concluded that the observed differences in learning outcomes are statistically significant. This evidence supports the superior effectiveness of the ChatGPT-based training approach, particularly for ADAS functions where operational complexity and non-intuitive physical placement may hinder learning. The slightly lower accuracy for the ACC function across both learning methods could be due to its more complex operation, requiring sequential button presses to engage and set speed, emulating manufacturers' real-time vehicle control systems (Mercedes-Benz, 2021; Toyota Motor Corporation, 2022).

Analysis Based on the Driving Experience

Driver experience is categorised as either a novice driver (one to three years), an intermediate driver (four to six years), or an experienced driver (over six years). This is consistent with (Commonwealth of Massachusetts, 2023; Robbins & Chapman, 2019). Our study investigated the relationship between driving experience and the

Table 6 Conventional group participant's response accuracy w.r.t their driving experience

Driving experience	AP-on response accuracy in %	AP-off response accuracy in %	LKA response accuracy in %	CA response accuracy in %	ACC response accuracy in %
Novice driver (1–3 years)	100	100	57	57	43
Intermediate driver (4–6 years)	100	100	70	70	54
Experienced driver (above 6 years)	100	100	84	89	74

Table 7 ChatGPT group participant's response accuracy w.r.t their driving experience

Driving experience	AP-On response accuracy in %	AP-Off response accuracy in %	LKA response accuracy in %	CA response accuracy in %	ACC response accuracy in %
Novice driver (1–3 years)	100	100	84	84	77
Intermediate driver (4–6 years)	100	100	90	100	82
Experienced driver (above 6 years)	100	100	100	100	88

proficiency with which participants engaged with ADAS. This is outlined in Tables 6 and 7 for the conventional and ChatGPT learning groups, respectively. Our findings concur with (Murtaza et al., 2023) that increased driving experience corresponds to improved performance in activating ADAS functions.

It is clear from Table 6 that novice drivers in the conventional group had lower accuracy rates (57% for LKA and CA and 43% for ACC), reflecting a learning curve associated with these functions. As experience increased, so did proficiency: intermediate drivers (four to six years) showed improved accuracy (70% for LKA and CA, and 54% for ACC), and experienced drivers (over six years) reached higher accuracy levels (84% for LKA, 89% for CA, and 74% for ACC).

For the ChatGPT group (Table 7), novice drivers started with higher accuracy rates of 84% for LKA and CA and 77% for ACC. This trend of improvement was more noticeable as drivers gained experience, with intermediate drivers reaching 90% in LKA and 100% in CA and experienced drivers achieving a perfect 100% accuracy in both LKA and CA, with ACC at 88%. The ACC function showed slightly lower accuracy in both learning methods. This may be because it is more complex to use, requiring sequential button presses to activate and set the speed, which mirrors the vehicle control system of the actual vehicle (Mercedes-Benz, 2021; Toyota Motor Corporation, 2022).

The observed differences in performance between the conventional and ChatGPT groups indicate that the method of instruction may influence how drivers learn and apply ADAS and AV functions. The ChatGPT group's results suggest that this educational methodology may offer a more effective foundation for understanding these complex systems, particularly for novice drivers. These insights support the integration of advanced instructional tools, such as ChatGPT, into driver training programs. This integration aims to enhance the utilisation of ADAS and potentially improve road safety for drivers at all experience levels.

Table 8 Conventional group participant's response accuracy w.r.t their ADAS frequency of use

Groups- division w.*r*.t ADAS function frequency of use	AP-on response accuracy in %	AP-off response accuracy in %	LKA response accuracy in %	CA response accuracy in %	ACC response accuracy in %
Never/Occasionally used	100	100	53	60	46
Intermittent user	100	100	76	76	58
Regular user	100	100	85	85	71

Table 9 CHATGPT group participant's response accuracy w.r.t their ADAS frequency of use

Groups- division w.*r*.t ADAS function frequency of use	AP-on response accuracy in %	AP-off response accuracy in %	LKA response accuracy in %	CA response accuracy in %	ACC response accuracy in %
Never/Occasionally used	100	100	76	84	69
Intermittent user	100	100	100	100	86
Regular user	100	100	100	100	91

Analysis Based on the Frequency of use of the ADAS Function

In a comprehensive analysis of driver engagement with ADAS functions, we observed a consistent positive relationship between the real-life frequency of ADAS utilisation and the accuracy of participant responses. Analysing the conventional training group more closely, individuals who had never/occasionally used ADAS functions (in their real-life) demonstrated initial competencies at 53% for LKA, 60% for CA, and 46% for ACC. Intermittent users of ADAS in their daily driving have experienced noticeable improvements. The accuracy rates for LKA and CA have increased to 76%, while the accuracy for ACC has risen to 58%. Regular ADAS users within this group improved further, achieving 85% precision in LKA and CA and 71% in ACC, as demonstrated in Table 8.

Participants in the ChatGPT educational group who never/occasionally used ADAS functions demonstrated high accuracy rates, 76% for LKA, 84% for CA, and 69% for ACC. The intermittent users showed swift learning, with perfect scores of 100% for LKA and CA and an impressive 86% for ACC. Regular users in this group upheld these high precision levels, consistently scoring 100% for LKA and CA and improving to 91% for ACC, as demonstrated in Table 9.

Our findings concur with the research of (Murtaza et al., 2023), supporting the concept that consistent real-world utilisation of ADAS is associated with enhanced driver performance. The comparison between conventional training and the ChatGPT-based approach highlights the potential influence of teaching methods on understanding and implementing ADAS features. Participants who were trained using ChatGPT demonstrated enhanced proficiency, suggesting that this method might provide benefits for learning complex automotive technologies. This is especially evident among new and occasional users, where the ChatGPT group's performance was relatively higher than the conventional group, suggesting ChatGPT offers a potentially more intuitive learning experience. These results support the integration of educational tools like ChatGPT into driver training curriculums, with the potential to increase the effectiveness of ADAS usage and improve road safety for drivers of different experience levels.

Analysis Based on the Participants' Familiarity with ChatGPT

Table 10 demonstrates a positive correlation between the frequency of ChatGPT usage and the improvement in ADAS function response accuracy. Those who use ChatGPT regularly achieved 100% accuracy in performing most functions, with a 92% performance in ACC. This suggests that consistent interaction with LLM-based tools can enhance one's ability to learn effectively.

Intermittent engagement with ChatGPT demonstrated improved performance outcomes. This is evidenced by a 100% accuracy rate in AP-On and AP-Off responses and a high degree of proficiency of 95% in LKA and CA functionalities. These findings suggest that periodic interactions with LLM-based platforms can substantially augment learning processes.

Individuals with minimal or no previous interaction with ChatGPT showed comparatively lower yet acceptable proficiency in activating ADAS functions. Specifically, in LKA and ACC functionalities, they achieved 75% and 63% accuracy, respectively, which, although satisfactory, are less than those observed in individuals who engaged with ChatGPT intermittently or regularly, as demonstrated in Table 10. These functions, located on the shifter module and not as intuitively accessible as those on the steering wheel, presented a steeper learning curve for these users. The challenge mentioned by participants, which we collected through feedback discussed in Sect. 5.3, reflects their initial struggles with navigating ChatGPT. They recommended incorporating visual aids in ChatGPT-focused training to help new users easily find and use ADAS functions. However, most participants reflected that ChatGPT training is engaging, enjoyable, and interactive. They appreciated that ChatGPT simplifies the information retrieval process, allowing users to concentrate more on learning and understanding ADAS and AV functions. This observation resonates well with (Dwivedi et al., 2023; Shoufan, 2023; Tlili et al., 2023). A noticeable correlation exists between regular ChatGPT usage and proficiency in handling ADAS and AV functions, emphasising the importance of integrating LLM-based tools in training for complex technological systems.

Table 10 ChatGPT group participants' response accuracy w.r.t their familiarity with ChatGPT

Groups- division w.*r*.t their familiarity with ChatGPT	ChatGPT learning group: No of participants in the group	AP-on response accuracy in %	AP-off response accuracy in %	LKA response accuracy in %	CA response accuracy in %	ACC response accuracy in %
Never/Occasionally used	8	100	100	75	87	63
Intermittent user	20	100	100	95	95	85
Regular user	12	100	100	100	100	92

Reaction Time Analysis

Data from the two groups, one trained with a conventional method and the other with ChatGPT was initially subjected to a normality analysis using Shapiro-Wilk's and Anderson-Darling's test. This preliminary analysis was performed to guide our decision in choosing the appropriate statistical test, parametric or non-parametric, depending on the normality of the data.

For the group that underwent training via the conventional methods, covering all ADAS functions (AP-On, AP-Off, LKA, CA, and ACC), it was verified through Shapiro-Wilk's and Anderson-Darling's tests that the collected data did not conform to a normal distribution. In contrast, data from participants trained through the ChatGPT system showed mixed results. The data adhered to a normal distribution for certain ADAS functions, namely AP-On and AP-Off, as confirmed by Shapiro-Wilk's and Anderson-Darling's tests. However, the data did not follow the normal distribution for the remaining functions, LKA, CA, and ACC. Consequently, conventional statistical measurements that assume normality, such as the analysis of variance (ANOVA) test, did not apply to this dataset. Given these observations, we determined that non-parametric testing was the appropriate statistical method for our analysis. We utilised the Mann-Whitney U test, a recommended non-parametric test for comparing two unrelated samples when the data are not normally distributed (Field, 2013). Therefore, the interpretation of our results was approached with due consideration of these analytical choices and the inherent attributes of the collected data.

The key performance indicator examined in this section is the reaction time required by the participants to activate/deactivate ADAS functions correctly. This metric was selected to evaluate and compare the efficacy of the two different instructional methodologies in imparting the participants' critical operational skills related to ADAS.

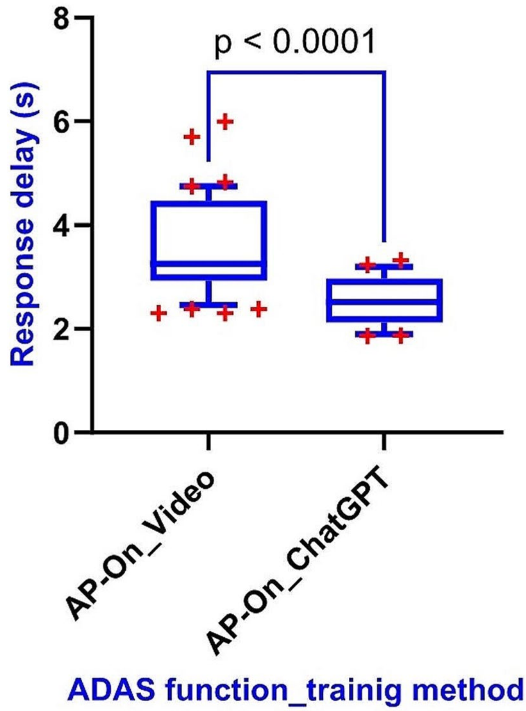
AP-On and AP-Off Functions

The box plots in Figs. 10 and 11 demonstrate unique patterns between participants trained using conventional methods and those trained using ChatGPT. It is clear from Fig. 10 (AP-On) group 1 that participants trained with the conventional method had a median response time recorded as 3.25 s. The span of response times for this group ranged from 2.30 to 6.00 s. This relatively wide range suggests a dispersion in data points, reflecting variability in the time taken by participants to press the correct button. The Standard Deviation (SD) for this group was calculated to be 0.92 s, underlining a higher variability in response times.

Conversely, participants in group 2, trained using ChatGPT, demonstrated a lower median response time of 2.51 s. Data from this group was more narrowly distributed, ranging from 1.80 to 3.32 s, and there was a significantly lower SD of 0.44 s. This tight range reflects greater consistency among participants, suggesting that they were quicker and more uniform in selecting the correct button. Outliers for both groups are marked with the "+" symbols on the plot.

In Fig. 11 (AP-Off condition), the median response time for the conventional method group increased slightly to 3.30 seconds, with responses scattered between

Fig. 10 AP-On reaction time comparison between Video vs. ChatGPT training methods. "+" shows the outliners



2.31 and 6.20 seconds. The SD for this group was calculated as 0.93 seconds. In contrast, the ChatGPT group in the AP-Off condition posted a median response time of 2.59 seconds. Their responses, ranging from 1.86 to 3.63 seconds and displaying a notably lower SD of 0.46 seconds, exhibited a tighter data clustering, indicative of more uniform performance among this cohort. Outliers for both groups are marked with the "+" symbols on the plot.

As presented in Table 2, the average accuracy for both AP-On and AP-Off conditions was recorded at 100% for all participants, irrespective of the training methods. However, the lower deviation and shorter response times of the ChatGPT groups suggest a better learning outcome of this approach.

A statistical comparison between the groups, employing the Mann-Whitney U test, was conducted for both AP-On and AP-Off conditions. The tests revealed statistically significant differences ($p<0.05$) between the two groups in both situations. Participants trained via ChatGPT consistently selected the correct response more rapidly than their conventionally trained counterparts. The highly significant p-value of less than 0.0001 provides compelling evidence for rejecting the null hypothesis, thereby validating the observed training advantage of the ChatGPT method in both autopilot conditions.

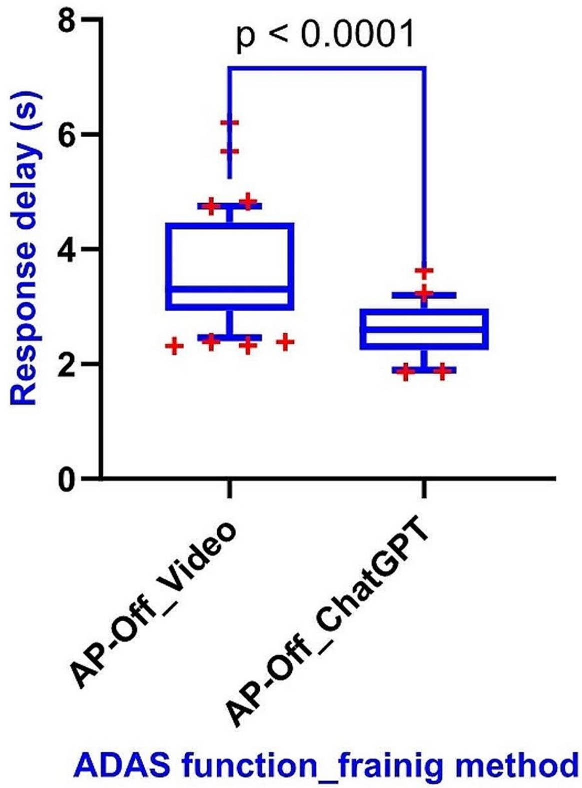


Fig. 11 AP-Off reaction time comparison between Video vs. ChatGPT training methods. "+" shows the outliners

LKA Function

The box plots in Fig. 12 demonstrate response times (in activating the LKA function) between participants trained through conventional methods and those trained using ChatGPT. It's noteworthy that, unlike the AP-On and AP-Off functions, the LKA function is situated on the shifter module instead of the steering wheel, a factor that may influence response times.

It can be observed from Fig. 12 that participants trained using conventional methods showed a median response time of 4.70 s. The response times for this group ranged from 2.52 to 8.84 s, with an SD of 1.92 s, highlighting a broad distribution in the time participants took to activate the correct function, indicating a wider variation in individual performance. In contrast, the group trained via ChatGPT had a lower median response time of 3.59 s. Their response times were tightly packed, ranging from 1.99 to 6.60 s, with a lower SD of 1.03 s. This data suggests a higher level of consistency among the participants in this group, implying they were both quicker and steadier in selecting the correct command on the shifter module.

Statistical comparisons between the two groups using the Mann-Whitney U test highlighted a significant difference in the response times ($p < 0.05$). The test produced a p-value of 0.0037, demonstrating strong evidence that the response times between the two groups were not the same. This underscores the superior efficiency of the

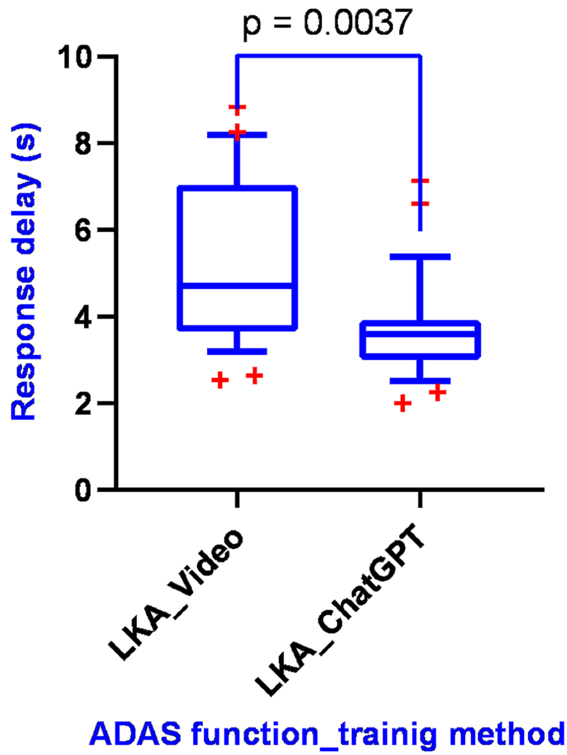


Fig. 12 LKA reaction time comparison between Video vs. ChatGPT training methods. "+" shows the outliners

ChatGPT training method, which showed significantly shorter response times than the conventional training method, even when interacting with the ADAS function located on the shifter module, which generally can take longer to reach.

CA Function

The box plots in Fig. 13 show the variability and central tendency of response times across both groups in activating the CA function. The group trained using the user manual recorded a median activation time of 4.40 s. In contrast, the ChatGPT-trained group demonstrated a faster median activation time of 3.18 s. This difference again suggests a quicker response time for participants trained through the ChatGPT model.

Analysing the spread of data offers additional insights. The group trained with the user manual showed a wide range of response times, extending from 2.02 to 8.54 s, with an SD of 1.64. This signifies a substantial variation in individual performance within this group. Conversely, the group trained with ChatGPT showed a more consistent range of response times, stretching from 2.10 to 6.40 s, and a lower SD of 0.98. This tighter spread suggests a greater consistency in performance among these participants.

The Mann-Whitney U test was used to ascertain the statistical significance of the differences observed. The resulting p-value was 0.0002, substantially below the commonly accepted significance level of 0.05. This indicates that the difference in median activation times between the two groups is statistically significant. These findings

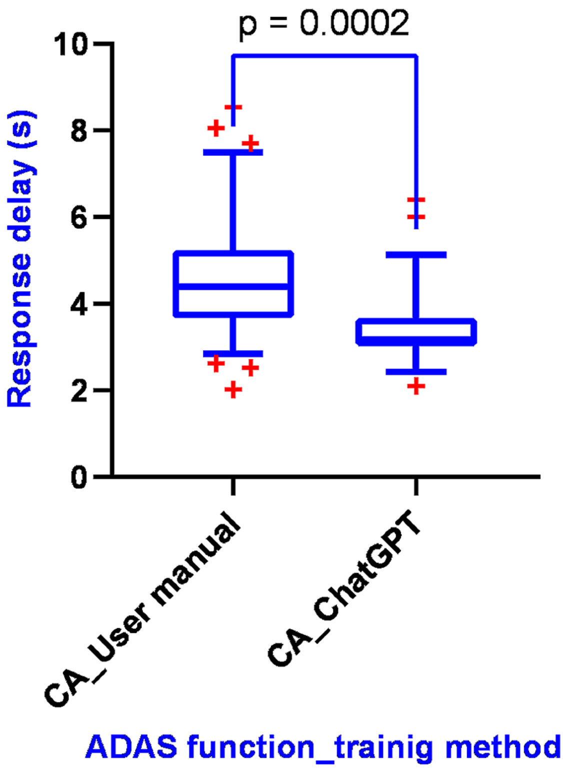


Fig. 13 CA reaction time comparison between User Manual vs. ChatGPT training methods. "+" shows the outliners

suggest that ChatGPT-based interactive training could lead to faster activation of the CA function than instruction methods relying on conventional user manuals.

ACC Function

The study further investigates the effectiveness of two distinct training approaches in instructing participants to operate the ACC function. This operation is slightly more complex, requiring the activation of two buttons: one to engage the ACC and a second to set the speed. This configuration was designed to emulate real-time vehicle control systems in models like those outlined in the (Mercedes-Benz, 2021; Toyota Motor Corporation, 2022).

The group trained via the conventional user manual displayed a median activation time of 6.95 s, with individual values ranging from 3.94 to 9.05 s. The SD for this group was 1.35, reflecting a substantial variation in individual performances. On the other hand, participants trained through the ChatGPT-based interactive user manual showed a notably more efficient performance. The median activation time for this group was significantly reduced to 4.91 s. Furthermore, the range of response times in this group was narrower, spanning from 3.04 to 7.90 s, with a lower SD of 1.007. This suggests a greater level of consistency in performance among the ChatGPT-trained participants.

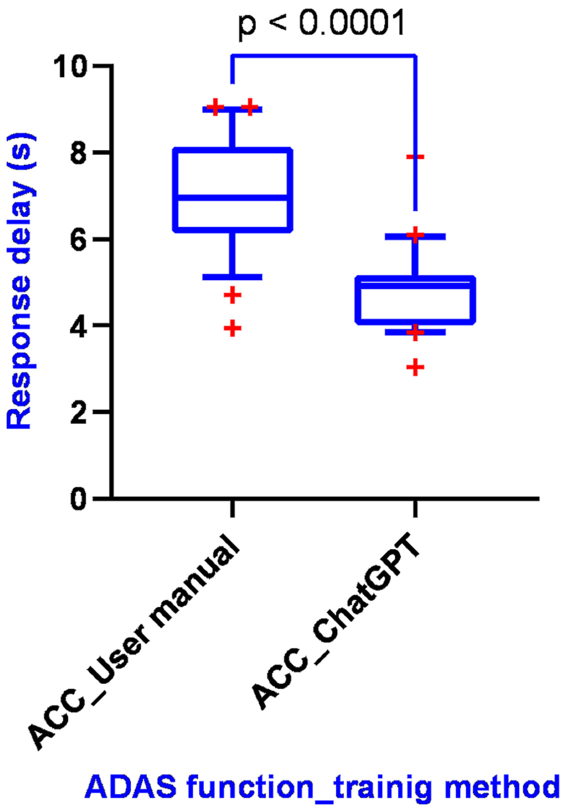


Fig. 14 ACC reaction time comparison between User Manual vs. ChatGPT training methods. "+" shows the outliners

A Mann-Whitney U test was performed to statistically assess the observed differences in median activation times between the two groups. As displayed in Fig. 14, the resulting p-value was less than 0.0001, significantly lower than the standard threshold of 0.05, indicating a statistically significant difference. These findings suggest that ChatGPT-based interactive user manual training is more effective in teaching participants to promptly activate the ACC function than the conventional user manual approach.

Discussion

Training Methods

This study aimed to determine whether the interactive and personalised instruction offered by ChatGPT could enhance learning efficiency and accuracy compared to conventional methods. Our key findings revealed that ChatGPT-based training led to faster activation times, increased consistency, and high accuracy across all examined ADAS functions.

The conventional learning process for ADAS functions is indeed diverse, involving demonstrations by sales agents at dealerships (Nandavar et al., 2023), video-based instruction (Merriman, Revell et al.,2023; Zahabi et al., 2020), and self-study through the vehicle's user manual (Boelhouwer et al., 2020; Greenwood et al., 2022). However, these methods have their limitations. For instance, due to time constraints and the complexity of ADAS features, dealership sales agents might not fully cover all available functionalities, leading to a knowledge gap (Murtaza et al., 2023). Furthermore, these agents may lack the necessary expertise to adequately explain the operation of ADAS (Abraham et al., 2017), further widening this informational gap. Additionally, user manuals, although detailed, often require a high level of existing knowledge and considerable time to be effectively used (Oviedo-Trespalacios et al., 2021). This is further compounded by the lack of interactive learning experience (Seel, 2011), which limits the effectiveness of these manuals.

A potential explanation for the superior performance of the ChatGPT-based training could be attributed to the interactive, responsive, and engaging nature of chatbot-based learning. Prior research (Dwivedi et al., 2023; Hirunyasiri et al., 2023) has suggested that applying LLMs, such as ChatGPT, enhances the educational experience by providing personalised instructions and immediate feedback. This feature potentially creates a more engaging learning environment that could facilitate quicker mastery of tasks and improved consistency in task performance.

The observed improvements of participants in learning outcomes can be further explained as follows: The interactive nature of chatbot learning infuses an element of gamification into the educational process. As (Wu & Yu, 2023) suggests, this aspect makes learning more enjoyable, thereby increasing user engagement. The 'game-like' interaction keeps the learners interested, providing a more dynamic, participatory, and effective learning experience. It is, therefore, reasonable to infer that the introduction of chatbots like ChatGPT into the learning paradigm has the potential to enrich the learning experience and motivate learners by making the process more enjoyable and engaging.

The growing complexity of ADAS demands innovative approaches to user education. Our study's findings suggest that AI-assisted tools like ChatGPT could potentially fill this gap, offering a more comprehensive and efficient learning experience than conventional video-based training or user manuals. The ChatGPT training methodology in our study demonstrated a high degree of accuracy across all tested ADAS functions.

The results presented in Table 2 reflect the significant benefits of using ChatGPT for learning ADAS functions. Notably, the ChatGPT group achieved an impressive accuracy range of 80–100% across all functions. Even when confronted with the complexities of ACC, the accuracy remained at a substantial 80%. In comparing these two instructional methodologies, the ChatGPT-based training consistently led to shorter median activation times and tighter data dispersion, indicating that participants were quicker and more consistent in activating the desired functions. Further research should be conducted to investigate the optimum number of training sessions to achieve even higher accuracy and to shorten the reaction time further in activating the ADAS or AV functions.

These results support the continued development and application of AI-based training methods like ChatGPT. However, further research is needed to corroborate these findings and explore the potential of such training methods across a broader spectrum of tasks and contexts. With technological advances and the increasing complexity of vehicle control systems, the need for effective and efficient training methods is evident. The findings of this study indicate that ChatGPT-based training could be a promising approach in this respect.

While promising, these findings should be interpreted cautiously, as other hidden factors could have influenced the outcomes. Future research should strive to replicate these findings across different settings and with larger participant samples to establish the generalisability of these results. The potential for extending the application of such AI-based training to other vehicle systems or even beyond the automotive industry is also worth investigating.

The Framework for Preparing a ChatGPT Prompt

In this work, we used a customised, ChatGPT-based interactive training platform to educate participants about the ADAS functions of a simulated AV. The principles we followed in this context ensured an effective training program and illustrated the broader potential of LLM across different industries.

The essence of effective training with ChatGPT lies in the level of specificity and context provided. We aimed to equip participants with comprehensive knowledge about each ADAS function. For instance, we detailed the concept, purpose, location, symbol, and procedure for activating/deactivating different functions. This meticulous approach eliminated any assumption of prior knowledge on the part of participants and reduced the possibility of confusion.

Ensuring the consistency of instruction delivery is another fundamental aspect of our method. We established a standard format for presenting information on ADAS functions. Additionally, we incorporated a proactive knowledge check mechanism into our training framework. This mechanism provides necessary information and asks a follow-up question to assess participant comprehension, thereby promoting iterative learning.

The constraints of AI systems, such as the absence of sustained memory in ChatGPT, necessitate regular reiteration of key instructions throughout user interactions (Open, 2023) and have been incorporated into the framework into some best practices, which ensure more accurate and contextually relevant responses. Furthermore, setting an output format personalised to each individual's learning experience can optimise the training efficacy.

Prompt engineers play a crucial role in enhancing user experiences with LLMs by guiding learners on how to effectively phrase and frame their questions or requests to utilise the LLM's full potential. We found that treating interactions with ChatGPT as two-way conversations, providing long-form questions, or sharing contextual narratives resulted in more precise and relevant responses.

Overall, our training method using a ChatGPT-based platform led to enhanced performance among participants compared to those trained using conventional methods. This was validated by the accuracy in activating the ADAS functions and reaction time in activating those functions. This suggests that the customised and interactive approach adopted for ChatGPT-based training was more effective in teaching participants about ADAS functions. Our study thus shows the immense potential of LLM-augmented approach training methods and their viability across various industries.

Subjective Opinions of Participants

The objective outcomes presented in Sect. 4 indicated improved accuracy and reaction times when using ChatGPT for ADAS training. To complement these findings, we collected the participants' subjective views to understand their personal experiences with the training process. These subjective experiences are crucial for interpreting the objective results, offering insights into the specific features of the ChatGPT training, which participants found beneficial. By examining these perspectives, we aim to uncover the factors contributing to their improved performance.

At the end of each participant's training, we collected their subjective opinions about the learning process with ChatGPT. Our study's qualitative analysis suggests that 90% of these opinions were positive. Participants highlighted ChatGPT as a dynamic and engaging environment ideal for learning ADAS functionalities. Notably, the system's interactivity and tailored approach were well-received, with learners appreciating the information's concise and direct delivery. Well-structured responses with highlighted keywords expedited the learning process and made it more enjoyable, integrating elements of gamification and motivation into the educational experience. These observed benefits are consistent with the studies (Dwivedi et al., 2023; Kasneci et al., 2023; Shoufan, 2023; Tlili et al., 2023). However, a minority of participants, approximately 8%, reported initial difficulties in interaction and a preference for additional visual aids to assist in the learning process. Despite these minor challenges, introducing the LLM into the training process positively influenced learning outcomes. In future research, we aim to conduct quantitative analysis, such as the NASA TLX (Task Load Index), to substantiate further the correlation between participant motivation, engagement, cognitive load, and learning efficiency. A key focus could be on examining the cognitive load during the learning process, particularly how AI-based learning compares to conventional methods in this aspect. A comparative study could also explore whether adopting LLM-based training tools can result in a reduced mental workload, thereby leading to improved performance and understanding of complex ADAS functionalities. Furthermore, conducting a study to determine if learning with LLM-augmented methodologies leads to quicker learning than conventional methods would be a valuable extension of our research.

Selected feedback from participants is as follows:

Positive feedback:

1. "Using ChatGPT has significantly enhanced my learning experience. It's particularly helpful for quick, direct responses to specific queries. The instructions given by ChatGPT were easy for me to follow and to locate the buttons. I like the output format in bullet points, and it allows me to step by step follow the instructions".
2. "Interacting with ChatGPT is enjoyable and makes learning easier because you can get direct answers to your questions. For example, when I asked about activating the Autopilot feature in a car, ChatGPT promptly directed me to its button

on the steering wheel. This immediate assistance saves time and makes learning efficient, enjoyable, and fun. I like that you do not need to search for the relevant information, ChatGPT does it for you".

Negative feedback:

1. "I found the learning process with ChatGPT a bit confusing since it didn't provide any visuals for the ADAS functions. I had to read the text and figure out the location of the button. It would be easier if ChatGPT included images of the ADAS functions".
2. "I had a bit of trouble learning with ChatGPT because it didn't show me any pictures of the ADAS buttons; I only got written explanations. In my opinion, images can help to make learning easier. Overall, I like learning with ChatGPT as I do not need to read all the details. You ask ChatGPT it filters for you".

The participants' reflections on using ChatGPT for ADAS training reveal its potential to enhance the educational experience. Their positive feedback underscores ChatGPT's intuitive interface and engaging interaction, suggesting that the AI-facilitated training could lead to more effective learning outcomes by making the process engaging and interactive. Additionally, the feedback highlights ChatGPT's role in streamlining the learning process by efficiently offloading the information retrieval task. ChatGPT enables participants to focus more on memorising and understanding the placement and use of ADAS features, as illustrated by participants' feedback for direct, step-by-step instructions and immediate, relevant answers.

However, the feedback also highlights areas for improvement, such as the need for visual aids to complement textual explanations. Recognising this, future research should consider the integration of multimedia into ChatGPT's reply. This advancement could potentially revolutionise the training experience by providing customised visual prompts alongside textual explanations, thereby accommodating the diverse audience for learning.

Extending LLM-Augmented Training to Other Industries

The previous sections demonstrated the framework's effectiveness in preparing a ChatGPT prompt to train drivers on ADAS and AV functions. Nonetheless, the proposed framework has broader applicability beyond the autonomous vehicle industry. For instance, it can be used in the manufacturing sector to facilitate the assembly of complex machinery. Consider an ordinary assembly task involving a jet engine in an aviation manufacturing setting. In this context, the proposed framework can effectively utilise a ChatGPT prompt for training technicians or engineers in assembling/maintaining a jet engine.

Goal and Scene Setting Start the interaction with ChatGPT by outlining the goals, methods, and conditions.

> Hello ChatGPT, in my next message, I will provide you with the contents of a detailed operational manual for assembling a jet engine. This manual includes comprehensive instructions on how to assemble the jet engine in a manufacturing setting. Instead of having the technical staff read the manual, you will interpret the contents and use an interactive, dialogue-based method to guide them through the assembly process [Approaches].
> Your goal is to ensure that the technicians or engineers can navigate through the complex assembly process of the jet engine efficiently, confidently, and safely after they have received this training. [Goals]
> Ensure that the guidance and instructions you provide to the technical staff are solely based on the specifics within the given manual and do not draw from generic knowledge or external resources [Conditions]

Then, the contents of the operational manual (in this case, the Jet engine assembly guideline) are provided to ChatGPT via the prompt. This approach assures that all participants receive consistent and accurate information about each task, avoiding potential confusion.

Personalise Responses The prompt engineer should customise the instructions and questions to match the audience's characteristics and task. In this scenario, where qualified engineers and technicians assemble a jet engine, it is important to instruct ChatGPT to use technical language and terminologies specific to the industry. The instructions should be detailed and precise, facilitating easy understanding and follow-through for the employees involved in the assembly process.

Providing Structured Instruction to ChatGPT Consistency is vital to helping ChatGPT to understand and respond correctly. It would be good to follow the standard process and procedure for each assembly step to tell ChatGPT about each part's assembly guidelines. For example, start with each part/instrument name, provide a detailed description, specify its physical location on the jet engine, describe any symbols or tools related to it, and finally, explain its execution procedure.

Proactive Knowledge Checking on Learners Train ChatGPT to answer a query with the corresponding answer and a relevant follow-up question to check the understanding of the engineers or technicians. For example, if a technician asks how to install a particular part, ChatGPT should present the steps and then ask a follow-up question: "Do you know where the part is located and what tools you need to use, or would you like me to explain in further detail?"

Reiterate Objectives and Conditions to ChatGPT Regularly It's crucial to remind ChatGPT every few prompts to adhere strictly to the instructions in the manual. This maintains the focus and accuracy of ChatGPT responses. Regular reinforcement of these guidelines is important, especially when a new conversation with ChatGPT is initiated or resumed after a pause.

Phrasing Output Format Guide ChatGPT in presenting its responses in various formats as needed. For instance, instruct ChatGPT to "*use detailed paragraphs for the description of each assembly step, use bullet points to list the steps for executing each step, and if the engineer or technician specifically requests information in a table format, provide it accordingly.*"

Conclusion

The comparative study on the efficacy of conventional instructional and LLM-augmented methods for teaching ADAS functions has yielded insightful findings. We found that using an LLM-based tool such as ChatGPT resulted in high accuracy and efficiency, which could enhance user education in advanced vehicular systems.

The implications of these findings are far-reaching, particularly for the automotive and transportation industry. ChatGPT could significantly bridge the knowledge gap concerning ADAS functions, leading to a higher adoption rate. Moreover, this novel instructional approach may reduce the number of incidents resulting from misuse or misunderstanding of these complex systems and thus yield better road safety.

For educators, our findings underline the potential of LLM in transforming user education, clearly indicating how this could shape future educational strategies. The effectiveness of ChatGPT, as shown in our research, suggests that LLM-based instructional tools could be included in considerations for future educational guidelines or instructional strategies surrounding the training for ADAS or other complex vehicular systems. There is a lack of comprehensive regulation or standardised platform for ADAS function training. Thus, these findings could be influential in initiating dialogues to develop such a framework.

Author Contributions Conceptualisation: Mohsin Murtaza, Chi-Tsun Cheng, Mohammad Fard, and John Zeleznikow. Methodology: Mohsin Murtaza, Chi-Tsun Cheng, Mohammad Fard, and John Zeleznikow. Software: Mohsin Murtaza. Validation: Mohsin Murtaza, Chi-Tsun Cheng, Mohammad Fard, and John Zeleznikow. Formal analysis: Mohsin Murtaza and Chi-Tsun Cheng. Resources: Chi-Tsun Cheng, Mohammad Fard, and John Zeleznikow. Writing—original draft preparation: Mohsin Murtaza and Chi-Tsun Cheng. Writing review and editing: Mohsin Murtaza, Chi-Tsun Cheng, Mohammad Fard, and John Zeleznikow. Visualisation: Mohsin Murtaza. Project administration: Mohsin Murtaza, Chi-Tsun Cheng, Mohammad Fard, and John Zeleznikow. All authors have read and agreed to the published version of the manuscript.

Funding Open Access funding enabled and organized by CAUL and its Member Institutions

Data Availability As part of RMIT University's human research ethical approval requirements and due to privacy concerns, supporting data is not available.

Declarations

Competing Interests The authors declare no competing interests. The authors have no relevant financial or non-financial interests to disclose.

References

Abdelghani, R., Wang, Y. H., Yuan, X., Wang, T., Lucas, P., Sauzéon, H., & Oudeyer, P. Y. (2023). GPT-3-Driven Pedagogical agents to train children's curious question-asking skills. *International Journal of Artificial Intelligence in Education*, 1–36. https://doi.org/10.1007/s40593-023-00340-7.

Abraham, H., Reimer, B., & Mehler, B. (2017). Advanced driver assistance systems (ADAS): A consideration of driver perceptions on training, usage & implementation. Proceedings of the Human Factors and Ergonomics Society Annual Meeting.

Agarwal, P. K., & Bain, P. M. (2019). *Powerful teaching: Unleash the science of learning*. Wiley. https://doi.org/10.1177/1045159519871920.

Al Kahf, S., Roux, B., Clerc, S., Bassehila, M., Lecomte, A., Moncomble, E., Alabadan, E., de Montmolin, N., Jablon, E., & François, E. (2023). Chatbot-based serious games: A useful tool for training medical students? A randomised controlled trial. *Plos One*, *18*(3), e0278673. https://doi.org/10.1371/journal.pone.0278673.

Aleven, V., Baraniuk, R., Brunskill, E., Crossley, S., Demszky, D., Fancsali, S., Gupta, S., Koedinger, K., Piech, C., & Ritter, S. (2023). Towards the Future of AI-Augmented Human Tutoring in Math Learning. [https://doi.org/10.1007/978-3-031-36336-8]. International Conference on Artificial Intelligence in Education, Tokyo, Japan.

Biswas, S. (2023). Prospective role of Chat GPT in the military: According to ChatGPT. *Qeios*. https://doi.org/10.32388/8WYYOD.

Boelhouwer, A., van den Beukel, A. P., van der Voort, M. C., Hottentot, C., de Wit, R. Q., & Martens, M. H. (2020). How are car buyers and car sellers currently informed about ADAS? An investigation among drivers and car sellers in the Netherlands. *Transportation Research Interdisciplinary Perspectives*, *4*, 100103. https://doi.org/10.1016/j.trip.2020.100103.

Brysbaert, M. (2019). How many words do we read per minute? A review and meta-analysis of reading rate. *Journal of Memory and Language*, *109*, 104047. https://doi.org/10.1016/j.jml.2019.104047.

Buch, S. V., Treschow, F. P., Svendsen, J. B., & Worm, B. S. (2014). Video-or text-based e-learning when teaching clinical procedures? A randomised controlled trial. *Advances in Medical Education and Practice*, 257–262. https://doi.org/10.2147/AMEP.S62473.

Chandra, S., Shirish, A., & Srivastava, S. C. (2022). To be or not to be… human? Theorising the role of human-like competencies in conversational artificial intelligence agents. *Journal of Management Information Systems*, *39*(4), 969–1005. https://doi.org/10.1080/07421222.2022.2127441.

Colabianchi, S., Bernabei, M., & Costantino, F. (2022). Chatbot for training and assisting operators in inspecting containers in seaports. *Transportation Research Procedia*, *64*, 6–13. https://doi.org/10.1016/j.trpro.2022.09.002.

Commonwealth of Massachusetts (2023). *Safe Driver Insurance Plan (SDIP) and your auto insurance policy*. Retrieved 25th Nov 2023 from https://www.mass.gov/topics/transportation.

Du, H., Teng, S., Chen, H., Ma, J., Wang, X., Gou, C., Li, B., Ma, S., Miao, Q., & Na, X. (2023). Chat with ChatGPT on intelligent vehicles: An IEEE tiv perspective. *IEEE Transactions on Intelligent Vehicles*. https://doi.org/10.1109/TIV.2023.3253281.

Dwivedi, Y. K., Kshetri, N., Hughes, L., Slade, E. L., Jeyaraj, A., Kar, A. K., Baabdullah, A. M., Koohang, A., Raghavan, V., & Ahuja, M. (2023). So what if ChatGPT wrote it? Multidisciplinary perspectives on opportunities, challenges and implications of generative conversational AI for research, practice and policy. *International Journal of Information Management*, *71*, 102642. https://doi.org/10.1016/j.ijinfomgt.2023.102642.

Ebnali, M., Hulme, K., Ebnali-Heidari, A., & Mazloumi, A. (2019). How does training effect users' attitudes and skills needed for highly automated driving? *Transportation Research Part F: Traffic Psychology and Behaviour*, *66*, 184–195. https://doi.org/10.1016/j.trf.2019.09.001.

Fagnant, D. J., & Kockelman, K. (2015). Preparing a nation for autonomous vehicles: Opportunities, barriers and policy recommendations. *Transportation Research Part A: Policy and Practice*, *77*, 167–181. https://doi.org/10.1016/j.tra.2015.04.003.

Fauzi, F., Tuhuteru, L., Sampe, F., Ausat, A. M. A., & Hatta, H. R. (2023). Analysing the role of ChatGPT in improving student productivity in higher education. *Journal on Education*, *5*(4), 14886–14891. https://doi.org/10.31004/joe.v5i4.2563.

Field, A. (2013). *Discovering statistics using IBM SPSS statistics*. Sage. https://uk.sagepub.com/en-gb/eur/request-inspection-copies.

Firat, M. (2023). How ChatGPT can transform autodidactic experiences and open education. *Department of Distance Education Open Education Faculty Anadolu Unive*. https://doi.org/10.31219/osf.io/9ge8m.

Gaspar, J. G., Carney, C., Shull, E., & Horrey, W. J. (2020). *The Impact of Driver's Mental Models of Advanced Vehicle Technologies on Safety and Performance [supporting datasets]*. A. F. f. T. Safety. https://rosap.ntl.bts.gov/view/dot/56626.

Greenwood, P. M., Lenneman, J. K., & Baldwin, C. L. (2022). Advanced driver assistance systems (ADAS): Demographics, preferred sources of information, and accuracy of ADAS knowledge. *Transportation Research Part F: Traffic Psychology and Behaviour*, *86*, 131–150. https://doi.org/10.1016/j.trf.2021.08.006.

Hirunyasiri, D., Thomas, D. R., Lin, J., Koedinger, K. R., & Aleven, V. (2023). Comparative Analysis of GPT-4 and Human Graders in Evaluating Praise Given to Students in Synthetic Dialogues. *arXiv preprint arXiv:2307.02018*. https://doi.org/10.48550/arXiv.2307.02018.

Ji, T. A., & Butterworth, S. (2019). Using Video and Paper–Based Educational Resources to Teach Common Surgical Techniques to Pre–Clerkship Medical Students: Results from a Simulation–Based Training Workshop. *UBC Medical Jornal*. https://doi.org/UBCMJ. 2019: 11.1.

Kasneci, E., Seßler, K., Küchemann, S., Bannert, M., Dementieva, D., Fischer, F., Gasser, U., Groh, G., Günnemann, S., & Hüllermeier, E. (2023). ChatGPT for good? On opportunities and challenges of large language models for education. *Learning and Individual Differences*, *103*, 102274. https://doi.org/10.1016/j.lindif.2023.102274.

Kay, J. (2023). Foundations for Human-AI teaming for self-regulated learning with explainable AI (XAI). *Computers in Human Behavior*, *107848*. https://doi.org/10.1016/j.chb.2023.107848.

Kolekar, S., de Winter, J., & Abbink, D. (2020). Human-like driving behaviour emerges from a risk-based driver model. *Nature communications, 11*(1), 4850. https://doi.org/10.1038/s41467-020-18353-4

Lehtonen, E., Airaksinen, J., Kanerva, K., Rissanen, A., Ränninranta, R., & Åberg, V. (2017). Game-based situation awareness training for child and adult cyclists. *Royal Society open Science*, *4*(3), 160823. https://doi.org/10.1098/rsos.160823.

Lehtonen, E., Perttula, P., Maasalo, I., Reuna, K., Kannisto, H., Puro, V., & Hirvonen, M. (2021). Learning game for improving forklift drivers' safety awareness. *Cognition Technology & work*, *23*, 743–753. https://doi.org/10.1007/s10111-020-00648.

Litman, T. (2017). *Autonomous vehicle implementation predictions*. Victoria Transport Policy Institute Victoria, Canada. https://www.vtpi.org/avip.pdf.

Lloyd, S. A., & Robertson, C. L. (2012). Screencast tutorials enhance student learning of statistics. *Teaching of Psychology*, *39*(1), 67–71. https://doi.org/10.1177/0098628311430640.

Lubkowski, S. D., Lewis, B. A., Gawron, V. J., Gaydos, T. L., Campbell, K. C., Kirkpatrick, S. A., Reagan, I. J., & Cicchino, J. B. (2021). Driver trust in and training for advanced driver assistance systems in Real-World driving. *Transportation Research Part F: Traffic Psychology and Behaviour, 81*, 540–556. https://doi.org/10.1016/j.trf.2021.07.003

MacFarland, T. W., Yates, J. M., MacFarland, T. W., & Yates, J. M. (2016). Mann–Whitney U test. *Introduction to nonparametric statistics for the biological sciences using R*, 103–132.

Mercedes-Benz (2021). *S-class operator's manual*. https://www.mbusa.com/en/owners/manuals.

Merriman, S. E., Revell, K. M., & Plant, K. L. (2023). Training for the safe activation of Automated vehicles matters: Revealing the benefits of online training to creating glaringly better mental models and behaviour. *Applied Ergonomics*, *112*, 104057. https://doi.org/10.1016/j.apergo.2023.104057.

Merriman, S. E., Plant, K. L., Revell, K. M., & Stanton, N. A. (2023a). A new approach for training needs analysis: A case study using an automated vehicle. *Applied Ergonomics*, *111*, 104014. https://doi.org/10.1016/j.apergo.2023.104014.

Murtaza, M., Cheng, C. T., Fard, M., & Zeleznikow, J. (2022a). The importance of transparency in naming conventions, designs, and operations of safety features: From modern ADAS to fully autonomous driving functions. *AI & Society*, 1–11. https://doi.org/10.1007/s00146-022-01442-x.

Murtaza, M., Cheng, C. T., Fard, M., & Zeleznikow, J. (2022b). Supporting Driver Training - from Vehicles with Advanced Driver Assistance Systems to Fully Autonomous Vehicles. Autonomous Vehicle Technology conference-APAC21, Melbourne, Australia.

Murtaza, M., Cheng, C. T., Fard, M., & Zeleznikow, J. (2023). Preparing drivers for the future: Evaluating the effects of Training on drivers' performance in an Autonomous Vehicle Landscape. *Transportation Research Part F: Traffic Psychology and Behaviour*. https://doi.org/10.1016/j.trf.2023.09.013.

Nandavar, S., Kaye, S. A., Senserrick, T., & Oviedo-Trespalacios, O. (2023). Exploring the factors influencing acquisition and learning experiences of cars fitted with advanced driver assistance systems (ADAS). *Transportation Research Part F: Traffic Psychology and Behaviour*, *94*, 341–352. https://doi.org/10.1016/j.trf.2023.02.006.

Nasir, A. R., & Bargstädt, H. J. (2017). An approach to develop video tutorials for construction tasks. *Procedia Engineering*, *196*, 1088–1097. https://doi.org/10.1016/j.proeng.2017.08.066.

Ng, D. T. K. (2022). Online lab design for aviation engineering students in higher education: A pilot study. *Interactive Learning Environments*, 1–18. https://doi.org/10.1080/10494820.2022.2034888.

Nick (2023). *Your AI Flight Instructor*. Part Time Pilot. Retrieved 8th June 2023 from https://parttimepilot.com/ai-flight-instructor-chat/.

Open, A. I. (2023). *The Challenges and Opportunities in Long-Term Memory for Language Models*. Open AI. Retrieved 25th June from https://community.openai.com/t/the-challenges-and-opportunities-in-long-term-memory-for-language-models/237178.

Oviedo-Trespalacios, O., Tichon, J., & Briant, O. (2021). Is a flick-through enough? A content analysis of Advanced driver Assistance systems (ADAS) user manuals. *Plos One*, *16*(6), e0252688. https://doi.org/10.1371/journal.pone.0252688.

Raf (2023). *What are tokens and how to count them?* Open AI. Retrieved 15 June from https://help.openai.com/en/articles/4936856-what-are-tokens-and-how-to-count-them#h_051eb08805.

Robbins, C., & Chapman, P. (2019). How does drivers' visual search change as a function of experience? A systematic review and meta-analysis. *Accident Analysis & Prevention*, *132*, 105266. https://doi.org/10.1016/j.aap.2019.105266.

SAE International. (2021). *SAE levels of driving Automation™ Refined for Clarity and International Audience*. SAE International. https://www.sae.org/blog/sae-j3016-update.

Salas, E., Bowers, C. A., & Rhodenizer, L. (1998). It is not how much you have but how you use it: Toward a rational use of simulation to support aviation training. *The International Journal of Aviation Psychology*, *8*(3), 197–208. https://doi.org/10.1207/s15327108ijap0803_2.

Sallam, M. (2023). ChatGPT utility in healthcare education, research, and practice: Systematic review on the promising perspectives and valid concerns. *Healthcare*, *11*(6), 887. https://doi.org/10.3390/healthcare11060887.

Seel, N. M. (2011). *Encyclopedia of the sciences of Learning*. Springer Science & Business Media. https://doi.org/10.1007/978-1-4419-1428-6.

Shoufan, A. (2023). Exploring students' perceptions of CHATGPT: Thematic analysis and Follow-Up survey. *Ieee Access: Practical Innovations, Open Solutions*. https://doi.org/10.1109/ACCESS.2023.3268224.

Smutny, P., & Schreiberova, P. (2020). Chatbots for learning: A review of educational chatbots for the Facebook Messenger. *Computers & Education*, *151*, 103862. https://doi.org/10.1016/j.compedu.2020.103862.

Susnjak, T. (2023). Beyond predictive learning analytics modelling and onto explainable Artificial Intelligence with Prescriptive Analytics and ChatGPT. *International Journal of Artificial Intelligence in Education*, 1–31. https://doi.org/10.1007/s40593-023-00336-3.

Thirunavukarasu, A. J., Ting, D. S. J., Elangovan, K., Gutierrez, L., Tan, T. F., & Ting, D. S. W. (2023). Large language models in medicine. *Nature Medicine*, *29*(8), 1930–1940. https://doi.org/10.1038/s41591-023-02448-8.

Tlili, A., Shehata, B., Adarkwah, M. A., Bozkurt, A., Hickey, D. T., Huang, R., & Agyemang, B. (2023). What if the devil is my guardian angel: ChatGPT as a case study of using chatbots in education. *Smart Learning Environments*, *10*(1), 15. https://doi.org/10.1186/s40561-023-00237-x.

Kluger Hybrid Owner's Manual. Toyota Motor Corporation, & Corporation, T. M. (2022). https://toyota-manuals.com.au/document/landing_page/kluger-hybrid-owners-manual-oct-22-current.

Van der Meij, H., & Van Der Meij, J. (2014). A comparison of paper-based and video tutorials for software learning. *Computers & Education*, *78*, 150–159. https://doi.org/10.1016/j.compedu.2014.06.003.

Wu, R., & Yu, Z. (2023). Do AI chatbots improve students learning outcomes? Evidence from a meta-analysis. *British Journal of Educational Technology*. https://doi.org/10.1111/bjet.13334.

Yuan, L. (2023). Where does AI-driven education, in the Chinese context and Beyond, go next? *International Journal of Artificial Intelligence in Education*, 1–11. https://doi.org/10.1007/s40593-023-00341-6.

Zahabi, M., Razak, A. M. A., Shortz, A. E., Mehta, R. K., & Manser, M. (2020). Evaluating advanced driver-assistance system trainings using driver performance, attention allocation, and neural efficiency measures. *Applied Ergonomics*, *84*, 103036. https://doi.org/10.1016/j.apergo.2019.103036.

Zheng, H., Mason, J. R., Classen, S., & Giang, W. C. (2023). Pilot study: Effect of roles and responsibility training on driver's use of adaptive cruise control between younger and older adults. *Transportation Research Part F: Traffic Psychology and Behaviour*, *94*, 53–66. https://doi.org/10.1016/j.trf.2023.01.023.

Zhou, L., Xu, S., & Qiu, Z. (2021). Intervening Construction Workers' Unsafe Behaviour with a Chatbot. [https://doi.org/10.1007/978-981-16-3587-8]. Proceedings of the 25th International Symposium on Advancement of Construction Management and Real Estate, Online.

Authors and Affiliations

Mohsin Murtaza[1] · **Chi-Tsun Cheng**[1] · **Mohammad Fard**[1] · **John Zeleznikow**[2]

✉ Mohsin Murtaza
mohsin.murtaza@rmit.edu.au

Chi-Tsun Cheng
ben.cheng@rmit.edu.au

Mohammad Fard
mohammad.fard@rmit.edu.au

John Zeleznikow
j.zeleznikow@latrobe.edu.au

[1] School of Engineering, STEM College, RMIT University, Melbourne, Australia

[2] Law and Technology Group, Law School, La Trobe University, Melbourne, Australia