Autonomous self-adaptive services for TRansformational personalised inclUsivenesS and resilience in mobiliTy

D4.2 Virtual assistant system and personalised interactions.v1

Lead beneficiary	CERTH	Lead author	Antonios Lalas, Konstantinos Votis
Reviewers	Lamberto Salvan (ALKE), Andrea Campos (CARITAS)		
Type	R	Dissemination	PU
Document version	V1.0	Due date	30/11/2025

Co-funded by the European Union

Project funded by
Schweizerische Eidgenossenschaft
Confédération suisse
Confederazione Svizzera
Confederaziun svizra
Swiss Confederation
Federal Department of Economic Affairs, Education and Research EAER
State Secretariat for Education, Research and Innovation SERI

한국산업기술진흥원
Korea Institute for Advancement of Technology

AutoTRUST project has received funding from CINEA under the European Union's Horizon Europe research and innovation programme under Grant Agreement No 101148123.

Project information

Project title	**Autonomous self-adaptive services for Transformational personalised inclUsivenesS and resilience in mobility**
Project acronym	**AutoTRUST**
Grant Agreement No	101148123
Type of action	Research and Innovation Actions (RIA)
Call	HORIZON-CL5-2023-D6-01
Topic	HORIZON-CL5-2023-D6-01-01 - User-centric development of vehicle technologies and solutions to optimise the on-board experience and ensure inclusiveness (CCAM Partnership)
Start date	1 May 2024
Duration	36 months

Document information

Associated WP	WP4
Associated task(s)	T4.3, T4.4
Main Author(s)	Antonios Lalas, Theofilos Christodoulou (CERTH)
Author(s)	Christos Basdeskis, Theofanis Siamatras, Panagiotis Lepentsiotis, Theoktisti Marinopoulou, Elefterios Blitsis, Dimitrios Tsiktsiris, Evangelos Athanasakis, Thomas Kopalidis, Stefanos Georgiadis, Angeliki Zacharaki, Konstantinos Votis (CERTH), Gerasimos Arvanitis (Avisense.AI), Lamberto Salvan (ALKE), Leonidas Katelaris (UCY), Faithful Abah (UNIGE), Giulia Consentino (UIA), Andrea Campos (CARITAS), Linda Mathe (SIEMENS)
Reviewers	Lamberto Salvan (ALKE), Andrea Campos (CARITAS)
Type	R – Document, report
Dissemination level	PU - Public
Due date	M19 (30/11/2025)
Submission date	30/11/2025

Document version history

Version	Date	Changes	Contributor (s)
v0.1	26/09/2025	Initial table of contents	Antonios Lalas, Theofilos Christodoulou (CERTH)
v0.2	07/11/2025	Contribution for section 1 and 2	Theofilos Christodoulou, Christos Basdekis, Theofanis Siamatras, Elefterios Blitsis (CERTH)
v0.3	14/11/2025	Contribution for section 3	Gerasimos Arvanitis (AviSense.AI), Lamberto Salvan (ALKE), Leonidas Katelaris (UCY), Faithful Abah (UNIGE), Giulia Consentino (UIA)
v0.4	24/11/2025	Contribution for section 3	Linda Mathe (SIEMENS)
v0.4.5	24/11/2025	Version ready for review	Angeliki Zacharaki, Konstantinos Votis (CERTH)
v0.5	25/11/2025	Initial review	Lamberto Salvan (ALKE)
v0.6	26/11/2025	Contribution for section 3	Andrea Campos (CARITAS)
v0.7	27/11/2025	Refinement based on review comments	Theofilos Christodoulou, Christos Basdekis, Theofanis Siamatras, Dimitrios Tsiktsiris (CERTH)
v0.8	27/11/2025	Review	Theofilos Christodoulou, Theofanis Siamatras, Theoktisti Marinopoulou (CERTH)
v0.9	27/11/2025	Review	Andrea Campos (CARITAS)
v0.9.5	28/11/2025	Final modifications	Theofilos Christodoulou, Christos Basdekis, Theofanis Siamatras (CERTH)
v0.9.8	28/11/2025	Refinement based on review comments	Antonios Lalas (CERTH) and CERTH team
v1.0	30/11/2025	Final version for submission	Antonios Lalas (CERTH)

Disclaimer

Copyright message

Contents

List of acronyms and abbreviations

Abbreviation	Description
AI	Artificial Intelligence
AIML	Artificial Intelligence Markup Language
AMS	Advanced Monitoring System
API	Application Programming Interface
AR	Augmented Reality
ASR	Automatic Speech Recognition
CAD	Computer-Aided Design
CCAM	Connected, Cooperative and Automated Mobility
CAV	Connected and Autonomous Vehicle
CoT	Chain-of-Thought
CPU	Central Processing Unit
DHM	Digital Human Model
ECA	Embodied Conversational Agent
FOV	Field of View
GPU	Graphics Processing Unit
HCI	Human–Computer Interaction
HMI	Human–Machine Interface
HRV	Heart Rate Variability
HUD	Head-Up Display
IK	Inverse Kinematics
IP	Internet Protocol
JS	JavaScript
KG	Knowledge Graph
LiDAR	Light Detection and Ranging
LLM	Large Language Model
LUBA	Loading on the Upper Body Assessment
ML	Machine Learning
NASA-TLX	NASA Task Load Index
NDR	Non-Driving-Related (Task)
OEM	Original Equipment Manufacturer
RAG	Retrieval-Augmented Generation
REBA	Rapid Entire Body Assessment
RIVA	NVIDIA Riva Speech Framework
ROS2	Robot Operating System 2
RTF *(contextual)*	Reaction Time Framework
RULA	Rapid Upper Limb Assessment
STT	Speech-to-Text
SUS	System Usability Scale

Abbreviation	Description
TTS	Text-to-Speech
TOR	Takeover Request
TOT	Takeover Time
TTC	Time-to-Collision
UEQ	User Experience Questionnaire
UI	User Interface
UCD	User-Centered Design
VA	Virtual Assistant
VAS	Virtual Assistant System
V2X	Vehicle-to-Everything
VR	Virtual Reality
WP	Work Package
XR	Extended Reality

List of figures

List of tables

Executive summary

The current deliverable (first version) derives from the work performed under Task 4.3 "Virtual Assistant framework and XAI content" and Task 4.4 "HMI, user apps and interfaces for personalised interactions" and constitutes the first integrated specification and implementation blueprint of the AutoTRUST Virtual Assistant System (VAS). Initially, the conceptual foundations, requirements, and design principles of the assistant are presented, including the evolution of conversational agents, the role of LLM-based assistants in mobility contexts, and the explainability and personalization needs. Next, the deliverable outlines the Virtual Assistant architecture, detailing the integration of LLM-based reasoning, multimodal context processing, ROS2 communication, and speech-based interaction modules. The connections with the Advanced Monitoring System (AMS) of WP3 are described through the definition of the semantic interfaces that allow the assistant to interpret high-level signals such as distraction, fatigue, emotion, acoustic events, and gesture patterns.

As a following step, the deliverable presents the planned mechanisms for context-aware reasoning, proactive support, and explainability, including the adaptation of symbolic, neuro-symbolic, and concept-based explanation methods. The speech pipeline (ASR/TTS) and the foundations for future multimodal extensions, such as avatar integration, are also described. Furthermore, the document details the design and development of the HMI and user-interface components required for personalised interactions, as foreseen in Task 4.4, including accessibility requirements, multimodal displays, preference transfer across devices, and mobile/web applications. Finally, the interfaces and data flows between modules are specified to support the alignment of WP3–WP4 developments and to facilitate the upcoming integration and validation activities under WP5.

It should be underlined that this is the first version of the Virtual Assistant system and personalised interaction framework. The VAS and HMI components will be further expanded, refined, and optimized and their final specifications and full implementation will be reported in the second version of the deliverable.

1. Introduction

This deliverable presents the first version of the Virtual Assistant System (VAS) and Personalised Interactions developed in WP4 of the AutoTRUST project. The document outlines the conceptual foundations, design methodology, technical specifications, and initial implementation of the Virtual Assistant, which serves as the main communication, reasoning, and personalization layer of the AutoTRUST ecosystem. The aim is to detail how the assistant transforms multimodal perception data, user profiles, and contextual information into adaptive, trustworthy, and inclusive interactions that enhance safety, comfort, and situational awareness in automated mobility.

The document addresses four main objectives:

a) to describe the methodological approach used to define the assistant's functional scope, requirements, and design principles, based on user needs, human factors, and system-level constraints.
b) to translate the system requirements, user needs, and multimodal perception outputs into concrete technical specifications for the VAS, including its reasoning pipeline, explainability mechanisms, personalization functions, and speech-based interaction modules.
c) to present the initial architecture of the assistant, including its core components, multimodal data flows, integration with the Advanced Monitoring System (AMS), and communication interfaces with the HMI and user applications.
d) to structure the development responsibilities and define the interfaces between modules in a way that supports coordinated implementation and efficient integration across the project.

The first part of the deliverable focuses on the concept and requirements of the Virtual Assistant, introducing the rationale for LLM-based interaction, the context-awareness needs of automated vehicles, and the explainability and personalization capabilities required for trustworthy user engagement. The subsequent sections present the assistant's architecture, including its LLM integration strategy, multimodal reasoning, ROS2-based communication, resource-optimized deployment on embedded hardware, and speech technologies for two-way dialogue. The document then details the design of the HMI and user applications that support personalised interactions, accessibility features, and cross-device preference continuity.

This version represents the initial implementation of the VAS. As the project progresses, further refinements, extensions, and validations will be performed in alignment with the developments of WP3 and the integration and evaluation activities in WP5. The final version of the VAS architecture and interaction framework will reflect the outcomes of these iterative improvements and will be presented in the second version of the deliverable.

1.1. Purpose and structure of the document

The purpose of the present document is to provide the first integrated description, specifications, and initial implementation details of VAS and the personalised interaction mechanisms developed in WP4 of the AutoTRUST project. It outlines the conceptual foundations, technical components, and interaction design principles that guide the development of the assistant as the central cognitive and communicative interface of the AutoTRUST ecosystem. The document also defines how multimodal perception, reasoning, explainability, personalization, and human–machine interaction elements come together to support safe, inclusive, and trustworthy user experiences.

Following the Introduction, which sets the stage for the document's purpose, audience, and its interconnections within the project's framework, the structure continues as follows:

2. **Section 2: Virtual Assistant Framework and XAI Content -** Presents the conceptual background, requirements, and technical underpinnings of the Virtual Assistant. This includes an overview of conversational agents, the rationale for LLM-based assistants in mobility, the role of explainability, and the assistant's position within the AutoTRUST architecture. The section then details the assistant's architecture, deployment strategy, multimodal context integration, proactive reasoning, and voice-based communication components.
3. **Section 3: HMI, user apps and interfaces for personalised interactions -** Describes the design principles, interaction requirements, and accessibility considerations that guide the creation of user-facing interfaces. This includes the multimodal HMI extensions, speech interfaces, dashboard functionalities, planned avatar integration, simulation-based validation approaches, and the role of mobile and web applications in enabling personalization, accessibility, and continuity of user preferences.
4. **Section 4: Conclusions -** Summarizes the main contributions of this first version of the Virtual Assistant System and outlines the next steps and expectations for the continued refinement, integration, and validation of the assistant and personalised interaction mechanisms in future project phases.

1.2. Indented Audience

The AutoTRUST "Virtual Assistant System and Personalised Interactions" deliverable is intended for public use as well as for the AutoTRUST consortium, including project partners, affiliated stakeholders, and external audiences interested in advancements in intelligent in-vehicle interaction systems. This document serves as a comprehensive reference describing the conceptual foundations, technical components, and interaction design principles of the VAS, together with the initial implementation of multimodal reasoning, explainability mechanisms, and user-oriented interface elements.

1.3. Interrelations

The AutoTRUST consortium integrates a multidisciplinary spectrum of competencies and resources from academia, industry, and research sectors, focusing on novel AI-leveraged self-adaptive framework for transformational personalised inclusiveness and resilience in CCAM. The project integrates a collaboration of fifteen partners from ten EU member states and associated countries (Switzerland, United States, and Korea), ensuring a broad representation for addressing security, privacy, well-being, health, and assistance, leading to enhanced inclusiveness, trust, and safety in the interaction between users and automated vehicles.

AutoTRUST is categorized as a "Research Innovation Action - RIA" project and is methodically segmented into 6 WPs, further subdivided into tasks. With partners contributing to multiple activities across various WPs, the structure ensures clarity in responsibilities and optimizes communication amongst the consortium's partners, boards, and committees. The interrelation framework within AutoTRUST offers smooth operation and collaborative innovation across the consortium, ensuring the interconnection of the diverse expertise from the various entities (i.e., Research Institutes, Universities, SMEs, and Large industries). This deliverable covers all activities related to the development of the VAS, including multimodal reasoning, explainability, speech-based communication, personalization mechanisms, and their integration with human–machine interfaces and user applications. It represents the main technical output of WP4 Tasks 4.3 and 4.4. The methodologies, algorithms, and processing pipelines documented here provide essential input to technical work packages such as WP3 and WP4, while also aligning with the requirements and specifications defined in earlier project stages. Furthermore, the outputs of this deliverable are designed to facilitate evaluation and validation processes in later phases of the project, thereby ensuring a seamless integration of the VAS within the broader AutoTRUST architecture.

Overall, D4.2 is positioned at the intersection of user needs, technological innovation, and ethical compliance, and serves as a key reference point for ongoing collaboration and knowledge exchange among AutoTRUST partners.

2. Virtual Assistant Framework and XAI Content

Section 2 provides the conceptual and technical foundation of the AutoTRUST VAS. It introduces the evolution of conversational agents, the motivation for adopting large language model (LLM)–based assistants in automated mobility, and the requirements that shape the system's functionality. The section details the architecture of the VAS, including its reasoning pipeline, multimodal integration, resource-optimized deployment, and explainability mechanisms. It also describes how the assistant processes semantic inputs from the perception system, produces context-aware natural-language feedback, and supports safe and inclusive interaction through speech-based communication. Together, these elements establish the core framework upon which the personalised and adaptable user experience of the AutoTRUST assistant is built.

2.1. Concept and requirements

2.1.1. Evolution of Conversational Agents and Chatbots

The development of chatbots and conversational agents has progressed through several distinct generations, shaped by advances in natural language processing (NLP), machine learning (ML), and artificial intelligence (AI). The field originated from early research on machine intelligence, notably Alan Turing's 1950 paper "Computing Machinery and Intelligence", which introduced the Turing Test as a measure of a system's ability to simulate human conversation.

The first generation of chatbots, spanning the 1950s–1990s, relied primarily on rule-based systems that used pattern matching and scripted dialogues to imitate conversation. Pioneering examples include ELIZA (1966) [1], which simulated a Rogerian psychotherapist through keyword-based pattern substitution, and PARRY (1972) [2], which emulated a paranoid patient using predefined logical structures. Later systems like ALICE (1995) [3], built on Artificial Intelligence Markup Language (AIML), formalized chatbot construction and introduced reusable rule sets, while SmarterChild (2001) [4] marked the integration of conversational agents into mainstream messaging platforms.

A major leap occurred in the 2010s with the rise of deep learning and neural NLP models, enabling conversational agents to process natural language at scale. This period saw the emergence of voice-activated assistants such as Apple Siri (2011), IBM Watson (2011), Amazon Alexa (2014), Microsoft Cortana (2014) and Google Assistant (2016), and Google Duplex (2018) [4]. These systems combined automatic speech recognition (ASR), intent classification, and dialogue management pipelines to perform task automation and deliver context-aware responses in real time.

Their introduction marked the shift from text-based to voice-based, multimodal conversational interaction, setting the foundation for intelligent virtual assistants in automotive environments.

The current era defined by LLMs such as GPT, Gemini and LLaMA models [5],[6],[7] has transformed conversational AI into a generative and adaptive technology. Unlike prior intent-based systems, LLMs can perform reasoning, summarization, and multi-turn dialogue maintenance without task-specific training, thanks to the Transformer architecture and large-scale pretraining. This paradigm shift underpins the emergence of LLM-based virtual assistants, capable of few-shot learning, open-domain reasoning, and personalised conversational flow. These technological milestones collectively trace the evolution from symbolic and rule-based systems to data-driven neural architectures, culminating in generative multimodal assistants that integrate perception, cognition, and communication.

2.1.2. LLM-Based Virtual Assistants in the Automotive Industry

The automotive sector is undergoing a paradigm shift as manufacturers increasingly embed Large Language Model (LLM)–based assistants to enhance human–vehicle interaction. These assistants move beyond deterministic, command-driven interfaces toward generative, context-aware conversational agents capable of understanding intent, sustaining multi-turn dialogue, and reasoning over multimodal contextual cues. This industrial trend mirrors the transition observed in academic research from rule-based natural-language pipelines to adaptive and empathetic AI architectures that integrate perception, reasoning, and affective modeling.

Recent deployments across major original-equipment manufacturers (OEMs) illustrate this transition. Mercedes-Benz leads the shift with its MBUX Virtual Assistant, developed in partnership with Google Cloud's Automotive AI Agent, which leverages generative-AI reasoning for semantic understanding and predictive personalization. The assistant interprets voice and visual cues to deliver natural conversations and anticipatory vehicle control recommendations [8]. Likewise, Volkswagen has integrated ChatGPT into its IDA voice assistant across the ID-series electric vehicles and new Golf and Tiguan models, enabling open-domain, knowledge-based dialogues that extend beyond infotainment to contextual driving support [9].

Other European OEMs are following similar trajectories. Volvo Cars announced adoption of Google's Gemini LLM within its vehicles equipped with Google Built-In, marking one of the first in-vehicle integrations of a fully generative conversational agent for both infotainment and driver assistance [10]. Stellantis Group, has partnered with SoundHound AI to deploy a multilingual assistant powered by the ChatGPT API, delivering context-aware and emotionally adaptive dialogues across multiple European markets [11] .

Collectively, these initiatives demonstrate that LLMs are transforming vehicle assistants into cognitive co-pilots capable of real-time reasoning, personalization, and emotion-sensitive interaction. Tasks once limited to voice command execution now include dynamic route optimization, contextual control of cabin and infotainment functions, and user-specific content generation, all conducted through fluid, human-like conversation. At the same time, these generative systems raise new research and regulatory challenges, chiefly reliability, explainability, and safety of AI-generated content, highlighting the need for Explainable AI (XAI), transparent decision layers, and multimodal context-fusion mechanisms.

2.1.3. The Role and Impact of Virtual Assistants in Automotive and Mobility Contexts

The emergence of voice-based and multimodal Virtual Assistants (VAs) in mobility systems is reshaping how people interact with vehicles and transport infrastructures. Beyond simple command execution, next-generation assistants increasingly support cognitive, emotional, and accessibility functions that affect driver safety, trust, and overall experience. The literature demonstrates that the value of these assistants lies not merely in technological sophistication but in their capacity to adapt to human needs, interpret multimodal context, and communicate transparently.

2.1.3.1. From Assistance to Companionship: The Cognitive and Emotional Role of Virtual Assistants

Modern VAs play an active role in regulating driver workload, stress, and vigilance. Huang et al. [12] showed that a ChatGPT-4–based assistant (“Driver Mate”) improved driver alertness in simulated long-distance driving by varying conversational complexity and frequency, with low-complexity/high-frequency interactions improving alertness. Likewise, Huber et al. [13] demonstrated that a voice-based assistant explaining in-vehicle warning signals significantly reduced self-reported stress and increased drivers’ perceived control and confidence. While not affect-adaptive, these findings highlight how well-designed conversational agents can enhance psychological comfort in safety-critical scenarios Collectively, these studies illustrate how conversational agents can support driver alertness and emotional stability during challenging situations.

Trust and acceptance depend heavily on personalization and perceived social intelligence. Liu et al. [14] identify trust and perceived usefulness as major determinants of willingness to use in-car voice assistants and derive design strategies, such as more user-centred, preference-aware interaction, to strengthen these factors. Sogemeier et al. [15] further observed that cultural expectations shape how users evaluate anthropomorphism and latency: German participants valued concise, efficient dialogue, whereas Chinese participants responded more positively to expressive, human-like agents. These findings suggest that the assistant's persona, tone, and rhythm must adapt dynamically to individual and cultural contexts.

Taken together, the literature frames the automotive VA not as a passive service interface but as a context-aware companion, one that blends functionality with empathy. This "co-driver" paradigm motivates the inclusion of affective computing and explainable dialogue management in AutoTRUST, enabling the assistant to reason about both situational and emotional states while preserving transparency of intent.

2.1.3.2. Context-Aware and Multimodal Interaction for Safe and Inclusive Mobility

Safety and accessibility in mobility depend on the assistant's ability to manage information across multiple sensory channels and to adapt interaction to different user groups. Studies reviewed by Chu and Huang [16] show that multimodal cues—particularly the combination of speech with visual information—enhance comprehension, reduce ambiguity, and support situational awareness when presented in a synchronized and contextually appropriate manner. The authors also identified six recurring dimensions of VA design tone, gender, anthropomorphism, trust, situational awareness, and accessibility, concluding that effective assistants must reason over both environmental and user states. Proactivity, when implemented transparently [17], enhances safety and engagement, provided the assistant explains its interventions and respects user control. These insights inform the AutoTRUST approach of context fusion: integrating signals from in-cabin monitoring (e.g., distraction, emotion, drowsiness) with conversational reasoning to trigger timely yet non-intrusive feedback.

Inclusivity constitutes another essential dimension of multimodal interaction. Studies in mobility accessibility demonstrate that voice-based interfaces empower users with age- or ability-related limitations. Bokolo [18] conducted a systematic review of intelligent conversational voice-assistants for older adults' mobility and derived design requirements indicating that older users benefit from clear, empathetic speech, simplified dialogue structures, and human-like guidance that supports independent, confident mobility in smart cities. Sangrar et al. [19] showed that conversational interfaces in autonomous shuttles improved perceived safety and independence among senior passengers. In parallel, accessibility reviews of public transit technologies confirm that auditory interaction mitigates barriers posed by small visual displays or complex menus. Collectively, these findings highlight the necessity of inclusive multimodality, a design philosophy ensuring that virtual assistants support diverse sensory and cognitive profiles across private and shared mobility.

2.1.3.3. Toward Explainable, Human-Centered Automotive AI

A unifying theme across the literature is the demand for explainable and transparent interaction. As assistants evolve from rule-based automation to autonomous reasoning entities, their decisions and utterances must remain interpretable to preserve user trust. Researchers emphasise that affective feedback, personalization, and proactive context-handling are only effective when coupled with mechanisms that clarify why the system acts as it does. In automotive contexts where misunderstandings can have safety implications, Explainable AI (XAI) principles are therefore essential.

2.1.4. Explainable AI Techniques for LLM-Based Virtual Assistants

2.1.4.1. Prompt-Based and Reasoning-Based XAI

Recent work shows that explainability can be embedded directly into the reasoning process of large language models, without relying on external model introspection. These approaches often referred to as LLM-native XAI leverage the generative and reasoning capabilities of transformer-based models to produce self-generated explanations, intermediate reasoning steps, or critiques aligned with human-understandable logic.

The most widely studied category is Chain-of-Thought (CoT) prompting, introduced by Wei et al. [20], which enables LLMs to reveal intermediate reasoning traces rather than producing a single opaque answer. CoT has been shown to improve both accuracy and transparency in tasks requiring multi-step reasoning, mathematical deduction, or logical inference. Follow-up work demonstrates variations such as Self-Consistency [21], where multiple reasoning paths are sampled and aggregated, improving robustness particularly for safety-critical applications. CoT belongs to the broader class of reasoning-focused prompting, where the model is guided to externalize its internal reasoning processes.

A related class of approaches uses self-explanation and self-critique prompting, where the LLM is explicitly asked to justify, verify, or critique its own output. This is explored in studies such as Madaan et al.'s Self-Refine [22] and Shinn et al.'s Reflexion [23], where the model iteratively improves its answer by generating explanations and reflective critiques. These methods shift XAI from post-hoc justification toward explanation-driven reasoning, where the explanation is produced as an integral step of the reasoning process.

2.1.4.2. Retrieval-Augmented and Knowledge-Grounded XAI

A second major paradigm in modern LLM explainability is retrieval-augmented generation (RAG), which improves both transparency and factual grounding by forcing the model to rely on external, verifiable evidence. In RAG pipelines [24][25], the LLM retrieves relevant documents such as manuals, policies, safety rules, or sensor descriptions and generates an answer conditioned on retrieved evidence. This ensures that explanations can cite or paraphrase explicit sources rather than hallucinating.

Recent LLM XAI surveys highlight retrieval-based grounding as an important mechanism for evidence-based explanations, since retrieved documents can be shown as justification. Zhao et al.'s survey of LLM explainability [26], categorizes retrieval-based grounding as a form of evidence-based explainability, since the retrieved documents can be presented to users as part of the justification. Similarly [27], identify RAG as an essential mechanism for "transparent LLM decision-making,". More advanced variants integrate knowledge graphs (KGs), using them for structured retrieval or enforcing consistency between model outputs and symbolic relations.

2.1.4.3. Model Editing, Internal Probes, and Multi-Level Explanations

A third category of LLM-specific explainability work focuses on interrogating and editing the internal representations of language models. Studies classify these methods under the utilization of explainability, where insights into model internals are leveraged to correct or control model behaviour. This category includes techniques such as causal tracing, knowledge editing (e.g., ROME, MEMIT) [27], fine-tuning-based editing approaches (e.g., gradient-based or supervised fine-tuning techniques such as MEND [28]) and neuron-/module-level interventions. These methods aim to identify or directly modify the internal components associated with particular factual associations or behaviours. Although their primary goal is often robustness, safety, or factual correction, they implicitly provide a form of explanation by demonstrating that "this behaviour can be identified or modified here".

More recently, Paes et al. propose MExGen [29], a framework for generating multi-level explanations for generative language models. Their method produces coarse-to-fine attributions, beginning with sentence-level contributions, followed by phrase-level and word-level analyses, to explain how different parts of the input influence generated text. The authors argue that existing explanation techniques designed for classification models do not readily extend to generative settings, especially when dealing with long inputs or multi-sentence outputs. Their results show that no single attribution method is sufficient; instead, effective explanations for generative models require layered, multi-granular analyses, combining segmentation, perturbation-based attribution, and structured aggregation of influence across levels.

2.1.5. Explainable AI Techniques for Convolution Neural Networks

Continuing from the explainability approaches discussed for large language models, CNN-based systems also require transparent methods to reveal how visual decisions are formed, particularly in high-risk domains such as in-cabin driver monitoring. While LLM-XAI focuses on reasoning traces, CNN explainability is fundamentally spatial, emphasising which image regions and learned features contribute to a model's output. In our current work, we specifically examine Grad-CAM to assess how our CNN models interpret in-cabin imagery and to validate that the classifiers attend relevant driver-related cues. This approach directly supports the explainability requirements of the D3.2 solutions, where transparent and trustworthy decision-making must be demonstrated.

Grad-CAM provides class-specific heatmaps that reveal the internal attention of the CNN by highlighting the most influential regions of the input image. This enables us to verify whether the model focuses on meaningful cues such as facial orientation, eye activity, hand placement, or gaze direction when predicting distraction states. Such visual validation is essential not only for model transparency but also for ensuring that the system adheres to expected behavior in safety-critical environments.

Beyond transparency, these visual explanations assist in diagnosing incorrect or suboptimal predictions. When a model misclassifies a distraction type, Grad-CAM allows us to identify whether the CNN relied on misleading or irrelevant visual features—for example, background objects, lighting variations, or reflections. These insights guide improvements in data collection, annotation, and model design, contributing directly to more robust and interpretable D3.2 deliverable outcomes.

Figure 1 presents two example Grad-CAM visualisations generated from our driver-distraction experiments. The left image corresponds to a texting-on-phone scenario, where the attention map correctly concentrates on the driver's hand region and the mobile device, indicating that the model relies on the expected cues when identifying this distraction. The right image illustrates a smoking scenario, with the heatmap focusing on the driver's hand-to-mouth area, capturing the characteristic gestures associated with smoking. These examples demonstrate how Grad-CAM effectively supports the analysis and validation of the CNN-based components developed for the D3.2 solutions, providing clear evidence of how the model forms its conclusions for different distraction types.

Figure 1: Example Grad-CAM Visualisations Highlighting Model Attention for Texting and Smoking Distractions

2.1.6. Description of the role of the Assistant module in the project

The Virtual Assistant (VA) module should function as the core cognitive and communicative interface of the AutoTRUST ecosystem, transforming the multimodal outputs of the Advanced Monitoring System (AMS) into intelligible, context-aware, and personalised interactions. Using high-level semantic signals generated in WP3, such as distraction, drowsiness, emotional state, gesture patterns, acoustic events, and environmental observations, the VA should interpret driver and passenger conditions and provide spoken feedback, warnings, and recommendations that enhance situational awareness and support safe mobility. It should act as the interpretative layer that converts technical sensor-derived insights into natural-language, human-centric messages, ensuring that complex perception and prediction processes are understandable, accessible, and actionable. Beyond safety cues, the VA should support convenience and inclusiveness by offering infotainment control, route guidance, schedule information, and adaptive communication tailored to each user's preferences and cognitive profile.

In addition to its real-time alerting capabilities, the Assistant module should extend to broader mobility-support functions that strengthen the continuity and quality of the user experience. Building on personalization and adaptation mechanisms, the VA should provide route and navigation information, monitors unexpected events affecting the user's journey, and communicate relevant updates such as traffic disruptions, changes in navigation planning, or recommended alternative paths. It can deliver public-transport-related information, including arrival and departure times, multimodal transfer guidance, or delay notifications, thus bridging the in-vehicle experience with the wider connected mobility ecosystem. The VA may also inform users about weather conditions and forecasted changes that could impact driving or trip planning, helping them make informed decisions. All these interactions should be governed by Explainable AI techniques, ensuring that the assistant can justify its suggestions, explain the reasoning behind alerts, and adapt its explanation style to different users' needs and abilities. Through this combination of safety-critical support, mobility-related assistance, personalization, and transparency, the VA serves as the central orchestrator of human-machine interaction within AutoTRUST, contributing to a trustworthy, inclusive, and proactive mobility experience.

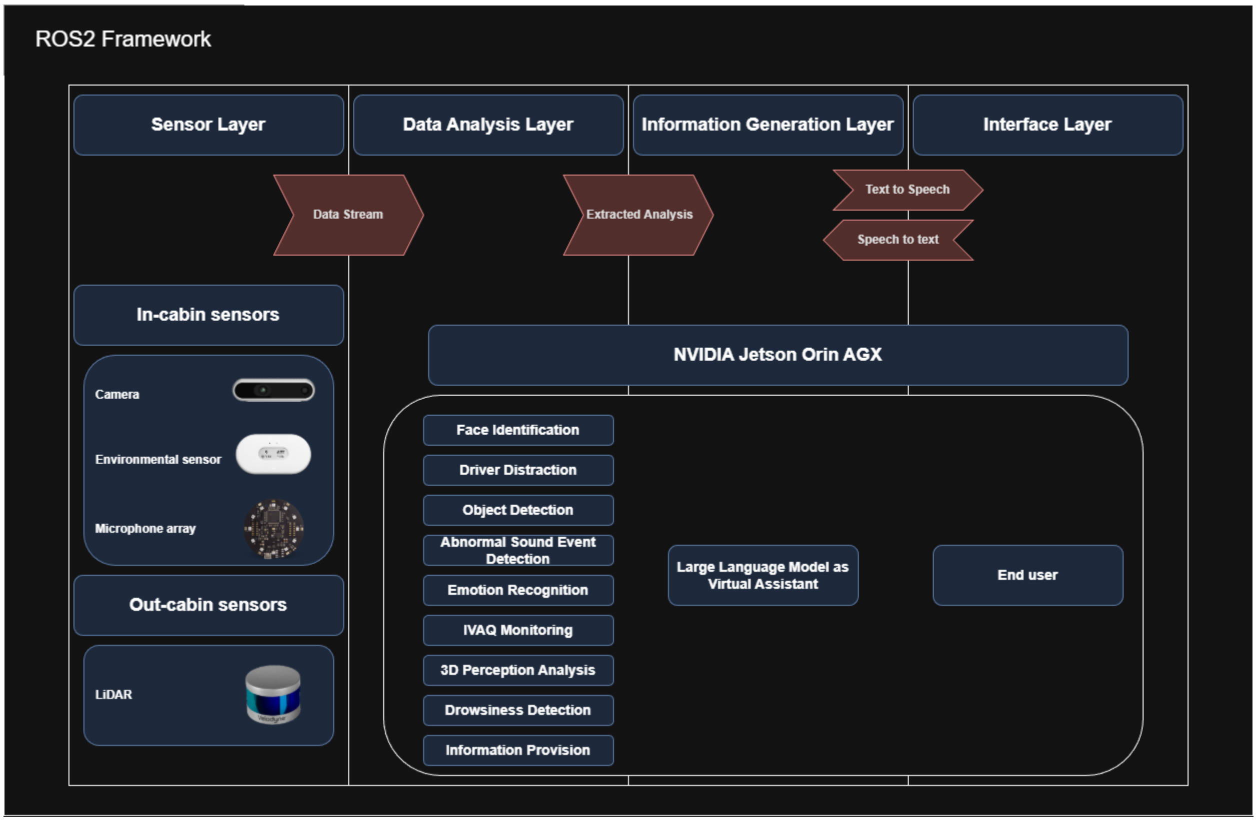

Figure 2: Conceptual Architecture of the Virtual Assistant

2.2. Virtual Assistant architecture and deployment

The virtual assistant (VA) acts as the primary cognitive and communication bridge between users and vehicles, enabling smooth and natural two-way interaction. Built as a ROS2 node, the VA combines multimodal sensory data (visual, auditory, and behavioral inputs) with a large language model (LLM). This configuration allows the system to read the driver's state, understand spoken commands, and generate adaptive responses that match the current situation. VA relies on perception-based reasoning and real-time conversation to enhance user engagement and situational awareness. It supports the safety and human-machine interaction goals defined in the WP3 Automated Mobility System (AMS) framework.

2.2.1. LLM integration

The virtual assistant's cognitive functions are primarily driven by a Large Language Model (LLM) responsible for contextual reasoning and natural language generation. This LLM is integrated into the ROS2 infrastructure via a specialized, lightweight interface. Critically, the model does not process raw sensor streams; instead, it subscribes to the processed, semantic outputs generated by dedicated perception modules, such as classifiers for emotion, distraction, drowsiness, and acoustic events. These aggregated and pre-interpreted signals are then contextualized by the LLM in accordance with a safety-oriented system prompt.

The inclusion of Automatic Speech Recognition (ASR) and Text-to-Speech (TTS) subsystems enables the LLM to engage in spoken dialogue. Through its cooperation with the ROS2 topics, the model remains continuously aware of the vehicle's operational context, serving as a modality-aware reasoning agent that autonomously maps sensor-derived events into concise and contextually appropriate communication.

2.2.2. Resource optimization

Developing advanced language and perception models on embedded vehicle systems, faces significant constraints in terms of memory consumption, processing power, and thermal management.

To address these constraints, the virtual assistant uses containerized development on NVIDIA Jetson hardware, incorporating model compression and efficient runtime techniques to reduce inference costs. System responsiveness is maintained through parallel and asynchronous processing, allowing perception, normalization, and messaging workflows to run simultaneously.

In addition, LLM inference is selectively scheduled to avoid unnecessary computations. The inclusion of preprocessing and temporal smoothing algorithms significantly reduces false positives and unnecessary model activations, thus saving computational resources and enhancing system resilience. This strategic synthesis of quantization-aware inference, container orchestration, and lightweight preprocessing enables the system to meet the demanding real-time performance requirements that characterize in-vehicle applications.

2.2.3. Connection with ROS2 and WP3 AMS modules

Serving as the main communication and reasoning element of WP3 Advanced Monitoring System (AMS), the virtual assistant is fully integrated into the ROS2-based framework. It connects the driver with the system's sensory functions by subscribing to ROS2 topics. These topics represent the semantic outputs from task-specific perception nodes that continuously monitor the driver's emotions, distractions, drowsiness, and auditory phenomena in the driver's environment. These modules convert raw sensor data from cameras and microphones into meaningful high-level signals that reflect the driver's condition and the environment. A vital pre-processing phase is performed to ensure the reliability of the interaction: perception data is improved over time before reaching the assistant. This system maintains a temporary memory of the driver's conditions and eliminates brief anomalies or incorrect measurements. The system evaluates the persistence and frequency of these signals (e.g., continuous fatigue, regular distraction) to ensure that the assistant's activations are triggered solely by reliable, significant patterns and not by misleading or temporary data points.

Moreover, a separate monitoring element consolidates these refined data streams to decide when an intervention is necessary. Every time a mix of signals suggests a possibly hazardous situation (like simultaneous distraction and drowsiness), this module creates a formatted message outlining the circumstances. The assistant then receives this event summary through ROS2 system. When received, the assistant's primary LLM clarifies this information, producing a relevant, natural language reply that is articulated via the TTS system.

This architecture establishes a unified pipeline from basic sensory perception to advanced cognitive interaction. The assistant functions as the smart interface of the AMS, responsible for converting intricate multimodal data into clear, user-friendly communication. Employing the ROS2 publish–subscribe model guarantees that the system remains modular, scalable, and in sync with all other components related to vehicle safety and monitoring.

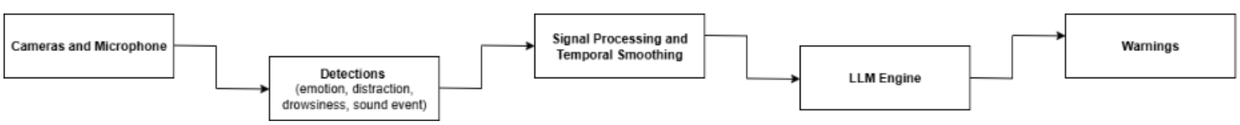

Figure 3: Data Flow of the Virtual Assistant

2.3. Multimodal interaction and Context-aware reasoning

2.3.1. Current integration of AMS modules (sound events, distraction, face ID, drowsiness, emotion)

The system uses several parallel processing pipelines originating from primary sensors. The front camera initiates two separate data streams from the raw video stream. One stream feeds an object detection node, which recognizes and publishes information about internal objects in the cabin, providing the assistant with information about the possible objects that can be found in the environment, such as a mobile phone or a laptop. At the same time, a face detection node isolates the driver's face and publishes a cropped image, creating a standardized input for a specialized "Driver" subsystem.

The "Driver" subsystem features a highly effective fan-out structure. The cropped facial data undergoes concurrent processing by three separate nodes: a driver identification node that recognizes the driver, an emotion analysis node that deduces the emotional state from facial expressions, and a drowsiness detection node that observes fatigue signs like eye closure. By allowing these resource-heavy artificial intelligence tasks to be executed concurrently, the system eliminates bottlenecks and supplies the assistant with precise information regarding the driver.

At the same time, the side camera records the driver's body and movements. The raw video is processed by a driver distraction node, which analyzes posture and can detect potential risky activities such as talking on the phone or drinking, and a gesture node, which recognizes specific hand movements that can serve as non-verbal commands. Complementing the visual data, a microphone records audio, which is processed by an audio detection node to recognize important events such as horn blaring, shouting, or even whether passengers are adults or children.

At the center of the system, the virtual assistant node acts as the main integration node and the brain of the architecture. It subscribes to all high-level data streams—including environmental objects, driver distraction, gestures, audio events, driver identity, emotions, and drowsiness — and synthesizes this information to construct a comprehensive, internal representation of the current context.

This associative synthesis allows the assistant to go beyond predetermined, programmed interactions. By combining multimodal signals in real time, the system can detect and respond to complex situations, such as driver fatigue or inattention, enabling adaptive interventions depending on the situation. While the current focus is on ensuring reliable synchronization of all data streams in real time, the architecture has been explicitly designed to support future enhancements to a higher level of reasoning. This will enable the assistant to infer predictive situations, assess risk, and adapt its feedback and dialogue strategies to the evolving context within the vehicle.

2.3.2. Planned context fusion and proactive support (signal prioritization, multi-signal reasoning, proactive suggestions)

The virtual assistant's next stage of development will focus on advanced environmental comprehension and predictive support. Currently, the system combines numerous data sources to generate an internal picture of the driver and his surroundings. The planned changes aim to go beyond simple reactive alerts, allowing the assistant to consider various signals and prioritize inputs based on urgency and relevancy.

To accomplish this, the assistant will use a hierarchical weighting mechanism that prioritizes different sensory inputs based on the scenario. For example, if fatigue persists, signals of driver drowsiness will take precedence over minor gestures, yet sudden loud sounds, such as the car horn, may momentarily overwhelm other observations. This ensures that the assistant concentrates on the most important occurrences, allowing them to respond more quickly and effectively.

Persistent support will also entail proactive communication. Using predictive driver behavior models and contextual indicators, the assistant will be able to provide advice before a potentially harmful situation arises. These interventions could range from benign statements like "Consider adjusting your driving," to critical safety alarms like "The vehicle in front is approaching too quickly, please slow down," all presented naturally by the system.

2.3.3. Explainability and trust (current direct alerts, planned multi-layer explanations, user adaptation)

The virtual assistant's primary goals are to increase driver confidence and ensure transparent operation. Currently, the system provides immediate alerts for critical situations such as fatigue, inattention, or dangerous noises. This approach is helpful for quick responses, but it doesn't give much insight into the assistant's motivations.

The multi-tiered explanations will be incorporated into the proposed improvements to make the assistant's decision-making process more understandable. When issuing a drowsiness warning, for example, the assistant may communicate: (i) the observed symptoms (such as prolonged eyelid closure or frequent yawning), (ii) consistent time trends, and (iii) the reasoning for the recommended action (such as exceeding the safety limit). The assistant aims to increase user trust and promote adherence to its recommendations by clearly relating interpreted information to decision-making processes.

Moreover, user adaptation mechanisms will be integrated to customize explanations and interaction styles according to personal preferences and understanding levels. Inexperienced drivers might get more thorough explanations, while seasoned users could get brief, actionable suggestions. This flexible method enhances user satisfaction and also consolidates trust in the system’s reliability and predictive abilities.

Ultimately, by incorporating explainable AI principles, context-sensitive reasoning, and adaptable interaction techniques, the virtual assistant seeks to be transparent and trustworthy, encouraging users to rely on its recommendations for safe and sensible in-car decisions.

2.3.4. Future integrations for Information provision

While the VA is currently designed to interpret multimodal driver-state information and support safety-critical communication, the architectural foundations developed within WP3 and WP4 open the possibility for future extensions that significantly enhance information provision during mobility.

A major future implication concerns integrating real-time arrival and departure updates for public transport, shuttles, and multimodal connections. As automated vehicles increasingly operate within interconnected transport networks, the VA could inform passengers about upcoming transfers, delays, service disruptions, or alternative options. Similarly, future implementations could incorporate dynamic route-change notifications, alerting users when road closures, congestion, or environmental factors affect their planned journey. This expands the VA from a reactive safety interface to a proactive mobility advisor that enhances planning, comfort, and user confidence.

The assistant may also extend its capabilities to include weather-aware mobility information, providing insights into how upcoming weather conditions may influence road safety, visibility, travel time, or the need for driving adjustments. For example, the VA could notify users about approaching storms, slippery surfaces, or reduced visibility, helping them anticipate risks and adapt accordingly. Such weather-informed dialogue is not part of the current AutoTRUST scope but aligns naturally with the project's emphasis on contextual, personalised, and explainable communication.

Overall, these potential functionalities illustrate how the VA could evolve beyond its immediate safety and personalization role toward becoming a comprehensive information provision platform, capable of supporting multimodal journey planning, environmental awareness, and adaptive guidance. While not included in the current project work plan, these extensions represent a promising direction for expanding the impact of the VA in future mobility systems.

2.4. Voice and speech-based communication

The virtual assistant uses multimodal speech technologies to create a seamless, natural, and context-aware interaction loop between the vehicle and its occupants. Speech works as both the primary input and output method, enabling hands-free communication that enhances driver safety and system accessibility. This conversational interface relies on two main components: an automatic speech recognition (ASR) module for processing spoken input and a text-to-speech (TTS) module for generating audible responses. Both systems integrate deeply with the central LLM reasoning engine, allowing the assistant to understand and articulate contextually appropriate messages.

2.4.1. Text-to-speech module outputs spoken alerts.

The text-to-speech (TTS) component turns the assistant's textual responses into natural-sounding audio. The current implementation uses Piper, an open-source neural TTS framework chosen for its optimization on embedded systems. Piper strikes a good balance between audio synthesis quality and computational efficiency, making it well-suited for resource-constrained hardware like the NVIDIA Jetson platform.

This module lets the assistant audibly deliver real-time, safety-critical warnings, driving recommendations, and situational updates. Audio-based delivery keeps the driver informed without pulling visual attention away from the road, which is vital for maintaining situational awareness. The synthesized voice output streams with minimal latency and high intelligibility to the vehicle's audio interface through the ROS2 communication layer.

Future development will focus on adapting vocal prosody and tone based on the driver's perceived emotional state. For instance, the system could use a calmer or more urgent delivery to match detected stress or fatigue.

2.4.2. Speech-to-text for two-way dialogue with the assistant

For speech input, the system incorporates an automatic speech recognition (ASR) pipeline powered by NVIDIA Riva. This component functions as a GPU-accelerated, server-based speech service that the assistant queries as a client. The client-server model handles high-performance processing in real time. The Riva service transcribes spoken language into text, which is then passed to the LLM for semantic interpretation and response generation.

A dedicated Riva server provides low-latency recognition with high accuracy, even in the challenging, noisy acoustic environment of a vehicle cabin. This allows the assistant to engage in natural, hands-free dialogue, supporting questions, confirmations, and conversation exchanges. The speech interface offers a safe and intuitive alternative to manual interaction, helping to minimise cognitive load and driver distraction.

The ASR pipeline monitors partial results to support real-time response. This allows the assistant to detect the user's intent before the sentence is complete, creating a more fluid conversational AI.

2.4.3. Avatar integration

While the current prototype focuses on voice, future work will integrate a 3D avatar, transforming the assistant into an Embodied Conversational Agent (ECA). This will leverage non-verbal cues—facial expressions, gaze, and gestures—to simulate more natural, face-to-face interactions

[30]. Recent research results shows that ECAs with non-verbal behaviors enhance social presence, engagement, and trust [31].

In in-vehicle applications, users respond better to sensitive feedback, such as fatigue or stress warnings, when agents display empathetic non-verbal cues [32]. Adding a visual avatar enables richer multimodal communication and a stronger connection with the user.

Integration will focus on synchronizing TTS audio with lightweight animation on an embedded platform, like NVIDIA Jetson. Generative AI models like Audio2Face-3D can create realistic facial animations from speech, producing blend-shapes compatible with real-time rendering [33]. Optimized for the limitations of edge computing, the assistant will evolve into a multimodal companion, capable of both speaking and expressing visual messages similar to those expressed by humans.

3. HMI, user apps and interfaces for personalised interactions

Section 3 focuses on the design and development of the Human–Machine Interfaces (HMIs), user applications, and interaction mechanisms that enable personalised and accessible communication between the Virtual Assistant System and end users. This section outlines the principles that guide the interface design, rooted in user-centered design, accessibility considerations, and human-factor insights, and describes the multimodal extensions required to support intuitive and inclusive interaction. It further presents the speech-based interface components, dashboard functionalities, validation methodologies, and mobile/web applications that complement the assistant's capabilities and ensure continuity of user preferences across devices. Collectively, these elements define the interaction layer through which users experience the assistant in everyday automated mobility contexts.

3.1. Ethics and the Responsible Research and Innovation

In accordance with the project's ethical framework, all personalisation features have been developed to increase user autonomy and control. Each adaptive function is optional and entirely subject to the user's decision, requiring informed consent in accordance with the GDPR (voluntary, specific and based on clear information) [34]. Users can modify or disable any adaptation at any time. To ensure transparency and understanding, the system provides clear feedback mechanisms, such as visual notifications or status reports, keeping users continuously informed about active personalisations [35],[36] In line with EU guidelines for trustworthy AI, the system emphasises human oversight, explainability, and user autonomy.

The system architecture must also incorporate transparency, reversibility, and accountability as essential requirements, ensuring that all data-driven adaptations can be understood, monitored, and adjusted safely by users.

Special attention is given to the needs of older people and users with reduced mobility. Interface elements and controls are designed according to ergonomic principles, with adjustable lighting, legible typography, and surfaces that minimise glare, ensuring visual comfort even with age-related changes [37]. Audible alerts use mid-range frequencies and allow for volume adjustment, making auditory information clear without causing discomfort.

All data-based adaptations include explanation and transparency mechanisms. Users can understand why a personalisation has occurred through simple summaries, notifications, or intuitive visual representations, and have the option to reverse or adjust any changes. These mechanisms promote trust, autonomy, and accountability, allowing users to maintain complete control over system interactions.

The system will incorporate fundamental ethical safeguards, ensuring that all adaptive decisions respect user autonomy and minimise potential risks. Continuous human oversight and the possibility of direct intervention ensure that no adaptive function operates in an opaque manner or produces unexpected effects without the user's knowledge and consent.

In the field of Connected and Automated Vehicles (CAVs), the approach follows the principles of Responsible Research and Innovation (RRI) [38] together with a structured set of ethical principles that guide technological development: non-maleficence, beneficence, human dignity, respect for autonomy, responsibility, justice, solidarity and inclusive deliberation. These principles ensure that the solutions developed improve safety, well-being and equitable access to mobility.

The project's ethical framework provides solid guidelines for decision-making, governance, and implementation within AutoTRUST, ensuring that the development of CAV systems occurs in a responsible, transparent, and socially aligned manner.

3.2. Design principles and requirements

3.2.1. Design principles and requirements for ICT products and services

This section defines the functional accessibility requirements for mobility applications intended for passengers, as well as for automated vehicle (AV) on-board passenger information systems. In general, the accessibility requirements defined in the European Norm EN 301 549 V3.2.1 (2021-03), "Accessibility requirements for ICT products and services" [39] shall apply both to mobile applications (Clause 11: Software, open functionality) and to passenger information systems (Clause 5.1: Closed functionality). Key requirements include appropriate colour contrast, visible focus indicators, the provision of alternative text for all user interface elements to support screen readers, magnification of content by up to 200%, sufficiently large interactive elements such as buttons, and suitable timing for user interaction.

In this context, closed functionality refers to functionality that is limited by characteristics that prevent a user from attaching, installing, or using assistive technology, whereas open functionality refers to functionality that supports access by assistive technology. Consequently, on-board passenger information systems shall provide built-in accessibility features, such as speech output, while mobility applications may rely on the assistive services commonly available on modern smartphones, provided that they properly support the accessibility services of the underlying software platforms. These assistive functions include screen readers, screen magnification, configurable gesture and touch interaction, and the use of sounds, vibration patterns, and adjustable visual parameters.

Even when a mobility application is designed to be accessible in principle, this alone is often insufficient for many persons with reduced mobility (PRMs), as such applications are typically used in demanding contexts and situations where users must simultaneously carry out other tasks, some of which may be time-critical or safety related. For example, users may need to locate the automated vehicle, orient themselves, and board the vehicle after receiving confirmation, often while carrying luggage or using assistive aids that occupy their hands. In such circumstances, there may be little or no opportunity to interact extensively with the application.

Furthermore, mobility applications present additional challenges for PRMs due to their complex functionality, the large volume of information displayed, the highly graphical nature of content such as maps, the high frequency of data updates (for example, real-time departure information), and operation in non-optimal environments with varying lighting conditions, noise levels, and weather effects.

3.2.2. Design principles and requirements for mobility applications for public transportation

Specific accessibility requirements and recommendations for mobility applications in public transportation are defined in DIN 13278:2022, "Smart mobility for persons with reduced mobility - Functional approaches" [40]. This standard has been published in both German and English and is currently in the process of being transposed into a European standard by CEN TC293 WG13.

The specific context in which mobility applications are used requires additional design considerations in order to optimize accessibility. In particular, it is recommended that a single application be used for all vehicles and user tasks, without requiring users to switch between multiple applications. Automated vehicles (AVs) should therefore provide open communication interfaces for mobility applications. Furthermore, users should not be required to register or log in merely to access information, while essential services such as stop requests or ticketing should be supported by simple and intuitive authentication mechanisms.

The application should support a high degree of individual configurability. Configuration settings should be protected against unintentional changes, and different user modes with predefined profiles for beginners, advanced users, and expert users should be available. Textual content and announcements should be brief, concise, and free of unnecessary filler words. Colours and contrast settings should be individually adjustable, and certain functions should be operable without requiring a change in input focus. Acoustic and haptic feedback, such as sounds and vibration patterns, should be used to indicate new events. The progress, successful completion, or failure of user interactions, as well as longer loading processes, should be clearly communicated through visual indicators and, where appropriate, through audio signals and vibration. User controls and navigation should be kept as simple as possible, while advanced functionalities should be accessible through submenus or optional expert views. Users should also be able to hide or rearrange interface elements according to their needs, for example by removing map displays that are not relevant to blind users. In time-critical situations, error messages should not interrupt or block ongoing interactions unless this is strictly necessary.

With regard to information presentation and services, all information and notifications should be delivered in a timely manner. Users should be able to repeat the most recent announcement without changing the current focus within the application. Information lists should be automatically sorted in real time according to relevance and temporal proximity. In cases where real-time data are not available, this limitation should be clearly communicated both visually and acoustically, and outdated information should not be displayed. It should be possible for users to hide unneeded information and services or to rearrange them based on personal preferences. The application should also provide straightforward options for applying temporary filters to reduce information overload and should allow the saving of favourites, recent destinations, filters, and predefined service requests.

Each user task should be supported by a dedicated view, and switching between views should be quick and intuitive. Typical views include vehicle location or “radar” displays, stop timetables, dynamic passenger information for a specific stop, route planning, and booking and ticketing functions. Wherever possible, operational tasks should be shifted to time periods in which the user is not under time pressure, for example by enabling advance booking and configuration of services.

At all times, safe travel must take priority over interaction with the application itself. In particular, when smartphones are used, blind and partially sighted users depend heavily on the consistent spatial layout of interface elements. Simple grid-based layouts with a limited number of elements in portrait orientation or linear lists between a header and a footer have proven to be especially easy to operate. Identical functions should always appear in the same location across all views of the application, and automatic reordering of interface elements should be avoided. For example, navigation back functions are typically expected in the top-left area, clocks in the top-right area, view-switching controls in the bottom row, and service request functions directly above it.

3.2.3. Design principles and requirements guiding the AutoTRUST project

The design of the Human–Machine Interface (HMI), user applications, and on-board interaction systems in AutoTRUST is guided by a comprehensive set of user-centered design principles derived from the multi-country engagement activities in WP2. These principles respond directly to the needs articulated by diverse user groups such as older adults, individuals with physical or sensory disabilities, visually impaired passengers, and general-population users, whose feedback consistently emphasised ergonomics, safety, inclusiveness, and clarity of interaction.

Across the focus groups, users expressed clear expectations regarding how information should be presented, how physical and digital elements should support autonomy, and how the system should behave in both routine and unexpected situations. These findings form the basis for the design requirements presented in this section, ensuring that all AutoTRUST interfaces provide equal access to mobility services, respect personal preferences, and support safe and comfortable travel for all user types.

The principles shaping AutoTRUST interfaces emerge from the user requirements analysis carried out for WP2, which revealed strong consensus across demographic groups regarding intuitive interaction, information transparency, and minimising cognitive load. Participants emphasised the need for systems that “explain themselves,” communicate predictably, and adjust to varied sensory or cognitive capacities. User-centeredness therefore requires that both user apps and in-vehicle interfaces adapt to the user rather than imposing a uniform interaction pattern. For older adults, this means simplified navigation and paced information flow. For visually impaired users, it requires consistent audiotactile guidance and elimination of overlapping sound layers. For persons with mobility impairments, accessibility begins with the physical layout and continues through reachable and logical UI elements.

Ergonomics and safety emerged as closely linked dimensions. Users frequently associated unclear instructions or sudden vehicle behaviour with feelings of insecurity. Therefore, the interfaces must not only be physically and visually accessible but must also reinforce perceived safety, for instance, by announcing upcoming manoeuvres or providing anchor-point confirmation for wheelchair users. Inclusiveness, understood as the capacity of the system to accommodate radically different needs without stigmatization, was another recurring theme across D2.3. Inclusive design requires multimodal communication pathways, adjustable sensory intensity, multilingual options, and respect for diverse privacy expectations. It also requires that no user be forced into modes of interaction that are inaccessible, overwhelming, or culturally inappropriate.

The user-facing mobile applications supporting mobility services act as the first point of contact for many passengers. Insights from D2.3 identified a set of accessibility needs that must be embedded from the early design phase of the mobility app. Participants with visual impairments stressed the importance of screen-reader compatibility and participants with visual impairments the importance of high-contrast layouts, zoom functionality (200%), simplified page structures, and consistent spatial organization.. Those with hearing or sensory sensitivities requested customizable notification modes, including vibration-only alerts and reduced visual clutter. For users with mild cognitive limitations or limited digital literacy (especially older adults) the app must avoid multi-layered menus, offer guided task flows, and provide explicit, step-by-step interaction sequences.

For users with mobility impairments, the app must present boarding information, ramp availability, and space allocation in advance, supporting self-reliant travel planning. The app should also integrate with the in-cabin interface to transfer user preferences (language, sensory preferences, accessibility mode) automatically upon boarding. This continuity was highlighted during the WP2 workshops as essential to reducing the cognitive effort imposed on passengers.

Finally, privacy-sensitive users represented by a consistent subset across groups require transparency regarding what data are collected and for what purposes, including the ability to opt out of non-essential sensing functions.

3.3. HMI and multimodal extensions

The human-machine interface (HMI) layer of the AutoTRUST framework serves as the main communication bridge between the driver and the vehicle's intelligent monitoring system. The multimodal interaction model combines visual, auditory, and (prospective) haptic channels to provide both situational awareness and physical interaction with the virtual assistant.

3.3.1. Dedicated HMI for driver assistance

Within AutoTRUST, a dedicated vehicle HMI is planned to provide a bridge between the vehicle control unit and the driver. Communication is achieved through visualisation of dedicated messages, as well as by providing a customized interface to best interact with the counterpart. Specific set-ups are created once the driver is recognized, maximizing interaction with a personalised user experience. The vehicle platform stores and remembers user preferences to provide consistent experiences across vehicles (for different users/activities).

The platform will be able to adapt the HMI display depending on the driver, supporting them in the event of cognitive disabilities that may limit use, such as colour blindness. This can be overcome by adapting the colour palette to the user's abilities. The platform will also support special operator activities.

The challenge will not be to exceed the user's limits only, but rather to create an interface specifically designed for the required activities (for example, a last-mile logistics operator will have an interface designed to optimise routes, loading and unloading, time management, and orders). The system will automatically adjust to the requirements dictated by real-time driver recognition.

Figure 4: Driver recognition system

Figure 5: HMI set-up including a dedicated monitor, on the right, for driver assistance

3.3.2. Text-to-speech - Speech-to-text integration

The multimodal HMI facilitates two-way voice communication between the driver and the virtual assistant. The speech-to-text (STT) system uses NVIDIA Riva, a powerful automatic speech recognition (ASR) engine designed for low-latency inference on embedded GPUs such as NVIDIA Jetson. The driver's voice commands are instantly converted to text and sent to the assistant's dialogue management system.

On the other hand, the text-to-speech (TTS) feature uses the Piper engine, a powerful open-source neural TTS system that generates natural speech locally on the Jetson platform. This ensures that verbal feedback and alerts can be generated independently of cloud services, maintaining privacy and immediacy.

The voice interface is fully integrated and embedded in a Docker-based setup, with ASR and TTS services running as ROS2 nodes that interact via standard message topics. The assistant can therefore understand the driver's instructions and provide voice alerts or confirmations. Two complementary interfaces have been developed for the virtual assistant:

- a conversation interface, where the user interacts through speech and receives verbal feedback
- an alert interface, which autonomously displays alerts and text messages derived from the perception units' "red flag" detections (e.g., distraction, drowsiness, unpleasant emotions).

Both run on top of ROS2 middleware, ensuring modularity and ease of integration with other functional units of the AutoTRUST system.

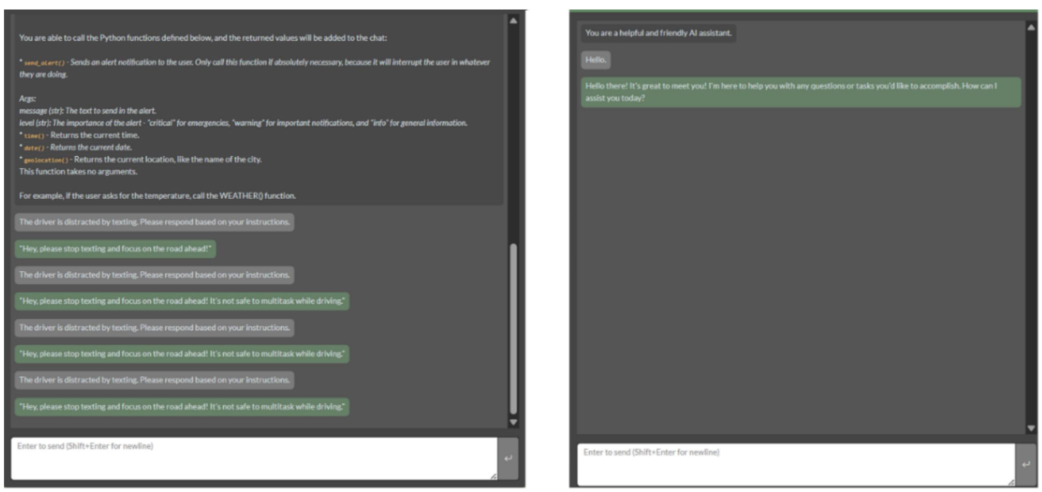

Figure 6: Virtual Assistant Interfaces

3.3.3. Dashboard (haptic interaction)

In addition to voice interaction, a web-based ROS2 dashboard has been developed to visualise the operational status of all perception and inference units. The dashboard, which has been implemented using Flask and JavaScript, dynamically subscribes to ROS2 topics via a backend node that aggregates and relays messages to the web interface.

Each perception unit (e.g., face recognition, emotion analysis, drowsiness detection) is represented by an interactive graphic element that changes color depending on its operating status: green indicates that the corresponding detection node is active and publishing data, while red indicates inactivity or lack of input.

Furthermore, the control panel includes real-time video streams from cameras viewing the driver and the cabin, spectrogram visualisations for the audio channel, and text summaries of the system's results (e.g., detected objects, recognized faces, emotion states). This visual layer provides a clear, concise picture of the system's situational awareness and can serve as a basis for future integration of haptic feedback, where tactile cues could accompany visual or auditory alerts for critical warnings.

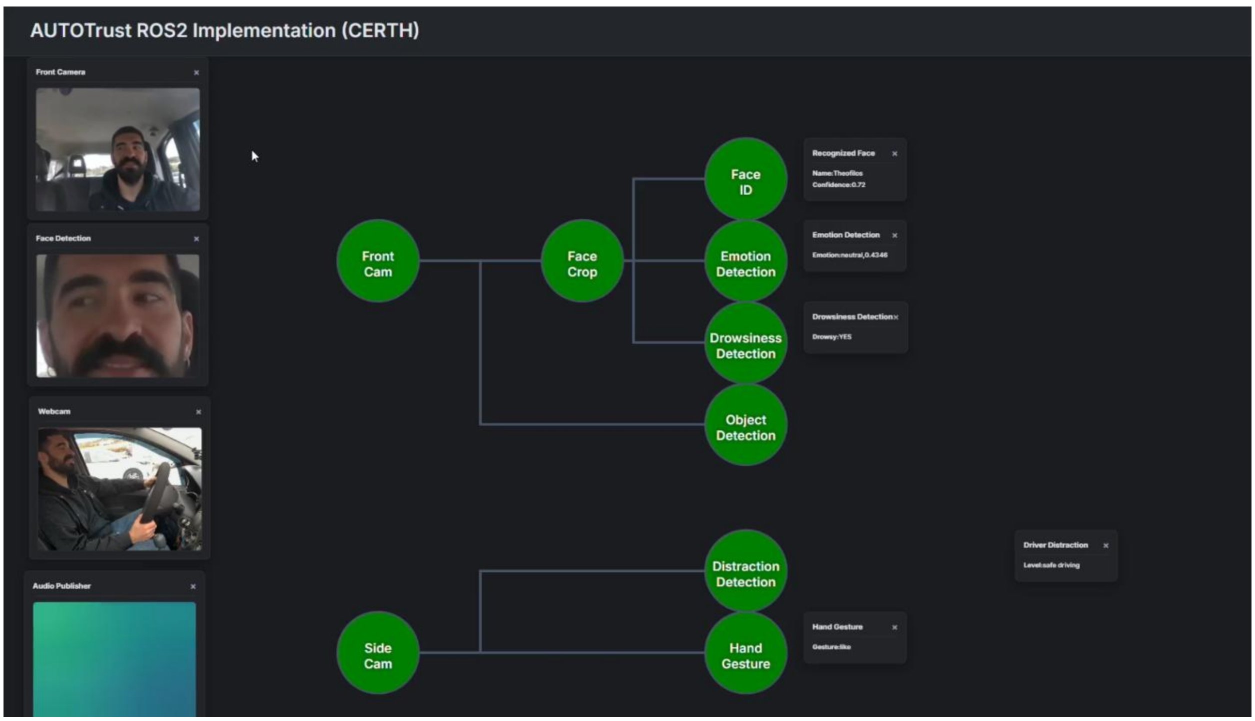

Figure 7: System Dashboard Interface

3.3.4. Avatar integration

An avatar element is planned as a future addition to the system, with the aim of providing a visual representation of the virtual assistant and enabling more attractive and intuitive interactions. The avatar will function as a meaningful interface element that will complement speech by conveying visual messages, such as attention or emotion, helping the driver to better clarify the assistant's responses.

This specific design and implementation approach is still under elaboration and will be determined in later stages of development. The emphasis will be on creating a lightweight, responsive solution that can work effectively alongside existing speech and dashboard interfaces. Once integrated, the avatar is expected to improve the overall user experience by promoting a more natural and human interaction between the driver and the system.

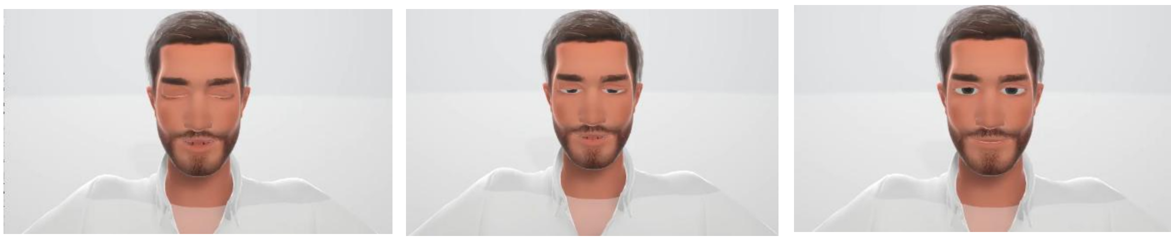

Figure 8: Concept Images for the Avatar Component

3.4. HMI simulations and validation approaches

This section presents the methodology, simulation toolkit, and validation strategy for the design and assessment of AR-based Human–Machine Interfaces (HMIs) within the AutoTRUST project. The objective is to ensure that future automated driving systems are ergonomic, trustworthy, intuitive, and safe, particularly during critical transitions between manual and automated control. AviSense.AI contributes by integrating biophysical simulations**,** XR-based HMI rendering**,** driver behavior modeling, and structured validation criteria for reaction time, perception, trust, and usability. Its role centers on developing human-centred, simulation-driven methodologies that i) analyse the ergonomic suitability of interface layouts, ii) model biophysical and postural interactions under different HMI configurations, iii) recreate complex road and weather conditions using XR/VR simulators, and iv) evaluate how drivers perceive, interpret, and react to system cues during semi-autonomous driving. To achieve these goals, AviSense.AI adopts an integrated workflow that combines:

- Digital human modelling and ergonomics assessment tools
- XR/VR driving simulators capable of replicating occlusions, environmental hazards, and multimodal alerts
- Biophysical and inverse-dynamics simulation models for posture and joint-load analysis
- Real-time behavioural monitoring (gaze, attention, drowsiness, head orientation)
- Trust and reaction-time assessment frameworks drawn from state-of-the-art human–automation interaction research
- Multimodal HMI prototyping**,** including AR overlays, HUDs, and XR content

The following subsections provide a detailed description of the simulation pipeline, ergonomic and biophysical analysis approaches, and the criteria used for validating HMI performance, safety, trust, and usability.

3.4.1. Develop biophysical/postural simulations for ergonomics optimization.

3.4.1.1. Objectives & relevance to AutoTRUST

A key objective of AutoTRUST is to ensure that future CAV systems enable safe, intuitive, and low-workload interaction between drivers and automated driving functions. Within this context, AviSense.AI develops a unified framework for biophysical modeling and postural simulation, allowing the analyses of how different HMI configurations affect driver comfort, posture, reachability, and situational awareness during normal driving operation.

Digital human models (DHMs) and ergonomic simulation methods have long been used to evaluate interior design, reachability, workload, and anthropometric accommodation in vehicle contexts [41]. DHMs, combined with inverse-kinematics and inverse-dynamics simulations, provide a systematic method for assessing:

- Driver posture under various seat, wheel, and HMI configurations [42]
- Reachability and effort when interacting with in-vehicle controls, XR interfaces, or other special operating devices (e.g., joysticks)
- Musculoskeletal loading, joint-torque accumulation, and ergonomic stress during prolonged use [43]
- Head and eye orientation relative to displays, augmented content, and situational-awareness cues [44]

These analyses are essential because CAVs introduce interaction patterns fundamentally different from conventional driving. Drivers in Level 3–4 systems frequently alternate between supervisory attention, non-driving-related tasks, and rapid re-engagement in dynamic environments, which directly affects posture, workload, and takeover performance. Poor ergonomic design, including suboptimal HMI placement, excessive reach distance, or forced head movements, can significantly compromise reaction time, decision quality, and trust in the automation. Integrating ergonomics early in the design process is very important, since in this way the simulations ensure that i) HMI layouts remain accessible and interpretable across diverse anthropometries [45], ii) visual and AR elements fall naturally within the driver's field of view, reducing unnecessary gaze shifts and cognitive effort, iii) physical workload (e.g., arm elevation, wrist strain, upper-body rotation) remains within biomechanical safety limits, iv) personalization strategies (e.g., adaptive HUD position, AR content scaling, seating recommendations) can be derived for each user type [46]. Ergonomics also plays a central role in shaping driver trust. When posture, visibility, and interface reach are well aligned, drivers experience greater confidence in their ability to monitor and intervene, along with reduced mental workload and clearer understanding of system behaviour. Properly designed ergonomic and HMI environments also help prevent overtrust and undertrust, ensuring alignment between perceived and actual system capabilities [47]. Finally, ergonomic optimization contributes directly to workload reduction.

3.4.1.2. Simulation framework and toolchain

The proposed framework enables detailed assessment of driver–vehicle interaction under different HMI layouts, postures, and operational scenarios, while supporting both functional and ergonomic optimization. The simulation environment begins with CAD-based interior models of the ALKE vehicle, representing the physical geometry of the cabin, seats, steering wheel, joysticks, dashboard components, and potential AR display locations. CAD-driven ergonomic analysis is widely used in vehicle design to evaluate reachability, control access, and HMI positioning with high geometric fidelity [41]. To explore population variability and ensure inclusivity, AviSense.AI generates digital avatars of diverse anthropometric profiles, systematically varying height, weight, gender, limb proportions, and torso dimensions. Digital human models and anthropometric scaling are commonly used to evaluate accommodation ranges and ergonomic suitability in automotive interiors [44]. This allows the simulation pipeline to model interactions for both the 5th-percentile female and 95th-percentile male, capturing extreme use cases and ensuring universal accessibility. These models provide quantitative indicators of biomechanical load when interacting with different HMIs, enabling early identification of unhealthy postures or excessive effort.

Figure 9: Apply human factors and ergonomics to redesign vehicle interiors, ensuring comfort and efficiency.

The framework includes motion modelling of driver interactions with various elements, including touchscreens, joysticks, steering interfaces, physical dials, and AR-based visual cues. This component builds upon established methods for analysing interactive reach motions, hand trajectories, gaze–head coordination, and seated interaction dynamics relevant to automotive contexts [46]. Interaction modelling is key for understanding how interface placement affects reach distance, posture deviation, muscular demand, and response latency. To evaluate physical strain, the workflow computes multi-joint energy accumulation and strain heatmaps, indicating regions of high biomechanical load across the upper limbs, spine, and neck during different tasks. Multi-joint stress visualisation and joint-energy mapping have proven effective for ergonomic risk detection, musculoskeletal workload evaluation, and design decision-making in vehicle HMIs [43], [47]. These heatmaps support both qualitative interpretation and quantitative safety assessment. The entire digital-human and biophysics framework is integrated with XR driving simulator, which recreates dynamic road scenarios, weather conditions, occlusions, sensor-limited visibility, and AR overlays. XR-based simulation environments have been shown to be effective tools for evaluating automotive HMIs, enabling safe and repeatable testing of driver behaviour under realistic cognitive load and environmental complexity [47], [48]. The coupling of DHM-based ergonomics with immersive simulation allows AutoTRUST to assess not only static posture but also behaviour during transitions, multimodal alerts, and temporal changes in situational awareness.

Figure 10: Driving aids, warnings, and alerts to enhance situational awareness and safety. Different weather and lighting conditions.

3.4.1.3. Posture scenarios and interaction tasks modeled

To ensure that HMI designs support safe, ergonomic, and efficient interaction, AviSense.AI simulates a wide range of posture scenarios and control tasks representative of real-world automated driving conditions. These simulations allow us to identify ergonomic risks, quantify movement effort, assess visibility and head orientation, and evaluate the feasibility of rapid transitions from automated to manual driving. The scenarios are grounded in established research on vehicle ergonomics, human–vehicle interaction, and driver behaviour modelling. A fundamental group of scenarios involves seat adjustments, steering-wheel reach, and pedal reach, which define the primary driving posture and strongly influence both physical comfort and driver performance. Studies show that seating geometry, reach distance, and steering-wheel/pedal configuration have direct impacts on musculoskeletal load, driving comfort, and reaction capability [41],[49]. These tasks help evaluate whether the cabin layout accommodates diverse anthropometries while maintaining safe biomechanics. Additional scenarios model joystick-based interactions made possible, once the updated ALKE CAD and control layout are integrated. Prior research highlights the importance of modelling such alternative control modalities to ensure accessibility and avoid upper-limb overload during prolonged use. Dashboard layout and control positioning substantially influence driver searching performance, head–arm movement, cognitive load, and safety-relevant distraction levels [44],[46]. Given the increasing use of augmented and mixed reality in automated vehicles, dedicated scenarios analyse gazing behaviour and head posture during AR use. Research demonstrates that the spatial placement of visual content significantly affects head tilt, gaze stability, situational awareness, and user comfort in XR-enhanced driving environments [41],[49]. Simulating these behaviours ensures that AR overlays fall within ergonomic and perceptual limits. Sudden re-engagement tasks, such as grabbing the wheel, touching a physical control, or responding to an AR-based takeover cue, are highly sensitive to posture, reachability, and visual access. Literature on automated driving consistently shows that physical layout, driver posture, and HMI configuration can meaningfully influence takeover time, movement quality, and safety margins during critical events [42].

Figure 11: Simulate avatars with different characteristics (e.g. height, weight, proportions) to represent a diverse user population

3.4.1.4. Ergonomic evaluation metrics and methods

To systematically evaluate the ergonomic quality of cabin layouts and HMI configurations, a comprehensive set of established ergonomic assessment tools, musculoskeletal modelling techniques, and internal biophysical indicators can be applied. These metrics quantify biomechanical load, posture risk, muscular effort, visibility constraints, and comfort across a wide range of driver anthropometries and use-case scenarios. The selection of metrics aligns with widely adopted methodologies in ergonomics, human–vehicle interaction, and musculoskeletal risk assessment. A primary metric used in the evaluation is the Rapid Upper Limb Assessment (RULA) score, a well-established method designed to assess the ergonomic risk associated with upper-limb, neck, trunk, and wrist postures during seated and reaching tasks [51]. RULA provides a standardized risk score based on joint angles and postural deviations, making it highly suitable for analysing touchscreen interactions, reach-to-control movements, steering adjustments, and AR-related head tilts. Complementing RULA, can also be considered REBA (Rapid Entire Body Assessment) and LUBA (Loading on the Upper Body Assessment) for evaluating whole-body postures and upper-body loading respectively. REBA is specifically tailored to unstable or dynamic postures in environments where frequent posture changes occur [52], making it useful for takeover manoeuvres requiring rapid reach-to-control. LUBA, on the other hand, focuses on static and semi-static upper-body loading and has been widely applied to evaluate seated work and in-vehicle interaction [53]. In addition to these standardized tools**,** joint torque thresholds and musculoskeletal load indicators derived from inverse-dynamics analysis, can be used, enabling more granular assessment of biomechanical risk. Musculoskeletal modelling approaches provide continuous estimates of joint moments, muscle activation levels, and accumulated exertion, variables shown to be critical for detecting ergonomic hazards in vehicle environments [54]. Others research involves the computation of joint-energy accumulation heatmaps, which visualise multi-joint mechanical load across task sequences. These heatmaps are derived from integrated kinetic and kinematic simulation outputs, representing cumulative joint stress over time. Similar biomechanical energy and strain-mapping techniques have been shown effective for identifying ergonomic risks and visualising workload distribution in human–machine configurations [43]. Finally, the system incorporates postural stability and comfort metrics, evaluating joint-angle ranges, posture variability, required trunk rotation, and head–neck alignment during HMI usage. Studies show that postural stability, reduced torso rotation, and minimal neck flexion are strongly associated with improved comfort and performance in automotive contexts. These metrics help quantify the ergonomic suitability of different AR placements and dashboard layouts.

3.4.1.5. Outputs & design recommendations

The generated simulation results translate into concrete design outputs and recommendations that guide the optimization of cabin geometry and HMI placement within AutoTRUST. These outputs ensure that vehicle interiors and AR-enabled interfaces are both ergonomically safe and accessible to a broad range of drivers. A key outcome of the biophysical simulation workflow is the derivation of seat and steering-wheel optimal positions tailored to different driver anthropometries. Research consistently shows that seat height, seat pan distance, lumbar support, and steering-wheel reach have strong effects on upper-limb comfort, trunk posture, and driver performance [41]. By analysing joint angles, postural load, and reachability across anthropometric extremes, AviSense.AI provides recommended adjustment ranges that minimise musculoskeletal strain while preserving rapid reach-to-control capability. Simulations further enable improvement of interior dimensions, including dashboard distance, control panel height, and spatial clearances, so that drivers of varying body sizes can achieve comfortable and biomechanically safe postures. These insights help refine interior layouts that accommodate diverse user groups without compromising safety or usability. Given the increased importance of mixed reality interfaces in automated driving, the simulation framework provides detailed recommendations for positioning AR/HMI elements, including optimal angles, heights, and field-of-view (FOV) parameters. Empirical findings indicate that the placement of digital overlays strongly affects head posture, gaze behaviour, visual workload, and situational awareness [44],[46]. Finally, the system will support a workflow for proposing personalised ergonomic adaptations, allowing cabin and HMI parameters to be tailored to individual drivers. This personalization strategy is grounded in research showing that adaptive seating geometry, dynamic display interfaces positioning, and user-specific control reach settings can substantially reduce biomechanical load and improve both comfort and trust in automated systems [50].

Figure 12: Create simulations to evaluate human interactions with HMIs

3.4.1.6. Next steps

A first priority is integrating the joystick CAD model into the ergonomic simulation environment. As joystick-based control interfaces become increasingly relevant in electric utility vehicles and future automated platforms, it is essential to evaluate reach distances, shoulder and forearm loading, and hand–wrist postures associated with joystick interaction. Existing research emphasises the need to assess alternative control layouts using digital human models to ensure safe and efficient manual override or fine-control tasks in automated vehicles [49],[54]. Incorporating the joystick geometry will therefore allow motion and musculoskeletal simulations to evaluate posture risk and accessibility for a broad range of drivers. Finally, AviSense.AI will focus on systematically linking ergonomic findings to HMI layout variations, enabling a data-driven approach for refining the AR intefaces positioning, dashboard configurations, control placement, and multimodal alert strategies. Literature shows that ergonomic indicators such as RULA scores, joint torque thresholds, and head/gaze alignment metrics should be directly incorporated into HMI design decisions to reduce distraction, cognitive load, and physical strain [46],[50]. This integration will allow us to evaluate how different display heights, angles, field-of-view constraints, and interaction distances affect both physiological comfort and task performance. By coupling ergonomic simulations with iterative HMI prototyping, the project can ensure that interface designs remain inclusive, safe, and optimally aligned with human capabilities.

3.4.2. Definition of validation criteria

3.4.2.1. Goals of the validation methodology

The purpose of the validation methodology is to ensure that the HMIs are safe, usable, trustworthy, and ergonomically optimized across diverse driving conditions and user profiles. The evaluation framework integrates behavioural, physiological, cognitive, and ergonomic indicators to provide a holistic assessment of driver–automation interaction. This approach is grounded in established research on automated driving, human–machine cooperation, and human factors engineering. A central objective is the measurement of **reaction time** and **takeover performance**, which are key safety indicators for Level 3–4 automated vehicles. Prior studies show that reaction time during takeover requests is directly influenced by interface design, driver posture, alert modality, and task engagement [21], [55]. By quantifying hands-on-wheel time, eyes-on-road recovery, and response latency, the validation framework identifies HMI configurations that support fast and stable transitions to manual control. Another major goal is the assessment of trust evolution, recognising that appropriate trust calibration, neither overtrust nor undertrust, is critical for safe automated driving. Trust in automation is known to dynamically vary based on system transparency, perceived reliability, and interface feedback [56]. The methodology also evaluates usability and cognitive load, both of which directly affect driver safety and acceptance. Research demonstrates that poor HMI layout, excessive information density, or unclear feedback mechanisms significantly increase mental workload and reduce task performance during supervisory automation [46]. Usability metrics (e.g., SUS, UEQ) and workload assessments (e.g., NASA-TLX) are therefore used to determine whether HMIs remain intuitive and cognitively manageable under different conditions. Increased pupil area and reduced gaze latency have been shown to correlate with heightened attention and improved takeover readiness in fog and low-visibility conditions [57]. These indicators allow evaluation of whether AR alerts effectively trigger perceptual recovery during transitions. Finally, the methodology assesses **situation awareness and ergonomics**, recognizing their combined importance for safe human–automation cooperation. Situation awareness is influenced by visual placement of AR elements, field-of-view constraints, and clarity of HMI cues. Ergonomic factors such as posture, reachability, and head–neck alignment also affect reaction quality and cognitive load. Overall, the goal of the validation methodology is to create a multi-dimensional, evidence-based evaluation framework that supports iterative, user-centred refinement of AutoTRUST HMIs and ensures their readiness for real-world automated driving applications.

Figure 13: Evaluate each avatar's interaction with the car interior to inform adjustments for improved ergonomic fit and comfort.

3.4.2.2. Reaction-time metrics and definitions

To evaluate the effectiveness and safety of HMI concepts within AutoTRUST, we will use a set of reaction-time (RT) and takeover-performance metrics grounded in the automated-driving literature. Reaction time is a critical determinant of driver safety during transitions from automated to manual control, and numerous studies have shown that HMI design, alert modality, environment complexity, and driver state significantly influence takeover behaviour [58]. A key performance indicator is **Takeover Time (TOT)**, defined as the elapsed time between the onset of a takeover request (TOR) and the driver's first meaningful response (e.g., wheel input, braking, or reaching for the control). TOT is widely used to assess takeover readiness, and research shows that TOT strongly correlates with workload, trust, and alert modality [59].Another critical measure is **hands-on-wheel time**, which reflects the time required for the driver to regain physical control of the steering wheel after a TOR. Studies indicate that this metric is sensitive to posture, HMI placement, and the cognitive engagement level immediately prior to the TOR [60]. Faster hands-on-wheel times generally correspond to safer intervention capability. To evaluate visual readiness, **eyes-on-road time** measures how quickly the driver shifts gaze back to the roadway or primary scene after receiving a TOR. Eye-tracking studies have shown that shorter gaze-recovery times are associated with higher situational awareness and improved manoeuvre quality [21]. This metric is especially relevant when evaluating AR alerts designed to guide gaze orientation. Safety margins during takeover are captured through **Time-to-Collision (TTC) reserve**, representing the available time before a potential collision at the moment the driver initiates control. TTC reserve has been used extensively to quantify takeover risk and to compare the safety of different HMI and warning systems. Higher TTC reserves indicate better preparedness and safer transitions. A related indicator is **lead-time adequacy**, which evaluates whether the TOR was presented with sufficient advance warning for the driver to reorient attention, regain control, and perform the required manoeuvre. Prior work shows that HMI-mediated alerts must be matched to road-scene complexity, as insufficient lead time is strongly linked to delayed or unstable response patterns [61]. The framework also includes metrics for **quality of takeover manoeuvre**, capturing lateral deviation, lane-keeping performance, braking smoothness, and steering stability. Research demonstrates that even with normalised reaction times, poor manoeuvre quality can occur if the driver is startled, cognitively overloaded, or visually misaligned during the takeover process. These qualitative metrics complement temporal measures, providing a more holistic assessment of safety. Finally, the validation process accounts for **the impact of alert modalities on reaction time**, drawing from evidence that **multimodal alerts**, combining visual, auditory, and tactile cues, produce significantly faster and more stable responses compared to unimodal signals. For instance, tactile cues have been shown to reduce reaction time in high-demand situations, while AR-based visual cues can reduce gaze-latency and improve perception of hazards.

3.4.2.3. Usability, cognitive load, and UX evaluation

To ensure that the developed HMIs are intuitive, easy to operate, and supportive of safe automated-driving interaction, a comprehensive set of usability, cognitive load, and **user-experience (UX)** evaluation instruments has to be applied. These measures capture subjective perceptions, cognitive workload, and overall interaction quality, complementing behavioural and physiological indicators. A standard component of the evaluation framework is the **System Usability Scale (SUS)**, a widely adopted 10-item questionnaire that provides a reliable estimate of perceived usability across diverse interface types [62]. SUS has been extensively applied in human–vehicle interaction studies to assess clarity, ease of learning, and interaction efficiency, making it suitable for evaluating both conventional dashboard elements and novel AR-based HMIs. **Experience Questionnaire (UEQ)** has been effectively used to evaluate automotive interfaces and provide deeper insight into the emotional and experiential qualities of HMI concepts, particularly those involving immersive or XR-based visualisations [63]. Cognitive workload is assessed using the

NASA Task Load Index (NASA-TLX), a validated multidimensional measure that evaluates mental, physical, and temporal demand, performance, effort, and frustration [64]. Numerous studies have shown that poorly designed automated-vehicle HMIs significantly increase mental workload, which can impair takeover readiness and reduce situational awareness [65].

Beyond specific instruments, the validation framework examines **HMI clarity and intuitiveness**, recognizing that interface transparency is essential for trust, safety, and low-effort interaction. Research consistently indicates that clear feedback pathways, consistent visual cues, and well-structured information architecture reduce cognitive load and prevent user confusion during automated-driving supervision [46]. These principles are applied to assess whether AR overlays, and multimodal cues communicate automation states clearly and avoid clutter. The evaluation also considers **perceived usefulness and user satisfaction**, which are key determinants of technology acceptance and trust in automation. Studies show that when drivers perceive HMIs as useful, supportive, and non-intrusive, they exhibit higher acceptance and improved reliance calibration, especially in automation contexts [66].

3.4.2.4. Trust measurement & trust dynamics

Trust is a critical factor in automated driving, influencing how drivers rely on system functions, respond to warnings, and engage during takeover events. The goal of a trust-evaluation framework is to assess both moment-to-moment trust dynamics and the overall alignment between perceived and actual system capability, also known as trust calibration. A core component of the framework is real-time trust estimation, inspired by Kalman-filter-based trust-dynamic models [67]. These models treat trust as a latent, continuously evolving cognitive state that can be inferred from observable behavioural and physiological indicators. By integrating real-time behavioural data with state-space models, the system can estimate trust levels without interrupting the driving task or requiring constant subjective input. **Gaze patterns** are evaluated as a key trust indicator, since prior research shows that attentive monitoring behaviour—shorter off-road glances, quicker gaze return to relevant cues, and stable fixation patterns—correlates with higher calibrated trust and better situational awareness. Conversely, erratic gaze behaviour or prolonged distraction can indicate undertrust, overtrust, or confusion, particularly during transitions from automated to manual control. Physiological responses such as **pupil dilation** are also incorporated as trust-relevant signals. Pupil area changes have been shown to correlate with cognitive effort, vigilance, and arousal during takeover transitions [21]. Increased pupil dilation following critical cues often reflects heightened attention and improved readiness, while flat or delayed dilation patterns may indicate low engagement or diminished trust. Behavioural performance indicators, such as **non-driving-related (NDR) task performance**, provide additional insight into trust dynamics. As demonstrated in automated-driving studies, high reliance on automation often leads to deeper NDR task engagement, while unstable or low trust reduces secondary-task performance because drivers allocate more cognitive resources to monitoring [56]. **System usage behaviours**, such as frequency of intervention, early manual override, or reluctance to engage automation, are also incorporated. Patterns of voluntary disengagement have been widely recognized as behavioural signatures of distrust or uncertainty, whereas failure to intervene when necessary, may signal overtrust. These behavioural cues support accurate mapping of how users balance reliance and oversight. Studies indicate that miscalibration, either undertrust or overtrust, can significantly impair safety, situational awareness, and takeover [68]. The validation framework therefore compares objective driving performance against both subjective and model-estimated trust levels to identify calibration issues. These scales complement real-time estimation by capturing the driver's holistic perception of the HMI and automation behaviour.

3.4.2.5. Perception arousal metrics

Perception arousal refers to the degree to which a driver's attentional and physiological system becomes activated in response to a takeover request or a potential hazard. The assessment of perception arousal is essential in automated driving, as drivers in supervisory roles often experience vigilance decrements, attentional disengagement, and delayed perceptual readiness. A primary metric is gaze duration reduction, which reflects how quickly a driver can shift visual attention back to the road environment after a takeover request. Research shows that shorter gaze duration on irrelevant stimuli and faster gaze redirection toward driving-relevant areas strongly correlate with improved takeover performance and higher perceptual alertness [12]. Gaze-transition metrics are therefore used to assess the effectiveness of visual and AR-based cueing strategies. Another key variable is pupil area expansion, which has been identified as a reliable indicator of increased cognitive arousal, attention, and perceptual readiness. In the Sustainability study [21] it was demonstrated that pupil dilation significantly increases immediately after a takeover request, particularly when drivers must rapidly shift from an NDR (Non-Driving-Related) task back to the roadway. This dilation is interpreted as heightened vigilance and improved processing of external hazards. Similar findings from controlled human-factors experiments confirm that pupillometry is a sensitive measure of transient attentional activation during driving [69],[70].

Perception arousal also includes assessing driver readiness to perceive hazards, often measured through gaze fixation patterns, detection times, and the time taken to visually acquire salient roadway elements (e.g., obstacles, vehicles, lane markings). It has been proved that drivers with higher arousal levels demonstrate shorter hazard-detection times and more consistent fixation behaviour immediately following a takeover request. These indicators provide insight into whether the HMI effectively helps the driver recover situational awareness under challenging visibility or weather conditions. Previous research in automated driving and attention studies has shown that decreased HRV and increased blink suppression correlate with heightened cognitive engagement and perceptual readiness [71]. While not always available in all test setups, these measures can complement gaze and pupil metrics.

3.4.2.6. AR information validation

AR technologies play a central role in next-generation automated-driving HMIs. The evaluation framework focuses on visibility, alignment accuracy, cognitive interpretability, and perceptual ergonomics, all of which are essential for ensuring that AR cues enhance, rather than hinder, driver performance. A primary component of the assessment is visibility and occlusion testing, which examines whether AR overlays remain visible under varying lighting, weather, and head-pose conditions. Research has shown that occluded or poorly visible AR cues significantly reduce interpretability and may impair situational awareness during automated driving [72], [73]. Through XR-based simulation, AR visibility is tested across conditions such as fog, glare, rain, and night driving, ensuring that digital overlays remain perceptually stable and unobstructed. A second critical factor is AR alignment accuracy, defined as the degree to which virtual elements are spatially consistent with real-world objects. Accurate alignment is essential for trust, comprehension, and hazard perception: misaligned AR cues can distort depth perception and create confusion in fast-moving traffic scenes. Prior studies show that incorrect AR registration can increase cognitive load and lead to delayed driver responses [74]. The validation framework also analyses driver comprehension speed, measuring how quickly users can interpret AR cues and translate them into actionable responses. It is well established that clear, minimally ambiguous AR cues significantly reduce information-search time and improve hazard-response behaviour. By conducting gaze-tracking and response-time analysis, it can be identified which AR designs enable rapid comprehension and which may require redesign to minimise cognitive delays. Another essential element is field-of-view (FOV) optimization, which ensures that AR elements fall within the driver's natural gaze envelope and do not force excessive head or eye movements. Studies indicate that improperly placed AR elements, too high, too low, or too peripheral, can degrade situational awareness and increase visual workload [75]. Finally, the avoidance of overload and visual clutter will be evaluated ensuring that AR systems communicate only the most relevant information in a clear and non-intrusive manner. Cluttered AR interfaces can lead to cognitive overload, inattentional blindness, and reduced hazard detection, particularly during takeover events or in visually complex environments [46],[76].

3.4.2.7. Simulator-based validation protocol

Simulator-based testing is widely recognized as a reliable method for studying human–automation interaction, allowing measurement of behaviour under hazardous or rare scenarios that would be unsafe to test on-road [60]. The VR/XR simulator supports a range of environmental conditions, including fog, rain, low-light/night driving, glare, and dynamically changing visibility. Prior work has shown that degraded visibility conditions significantly affect hazard detection, reaction time, and gaze recovery patterns, making them essential for testing AR-enhanced situational-awareness HMIs. By simulating such environments, the platform enables evaluation of whether AR cues improve perceptual performance during low-visibility takeover events. The validation protocol includes dynamic traffic elements such as crossing pedestrians, cyclists, and occluded vehicles emerging from behind obstacles. Occlusion scenarios are known to be particularly challenging for drivers and automated systems alike and are therefore critical for evaluating how well AR interfaces communicate hidden hazards [21]. These elements allow testing of HMI designs under realistic, high-demand traffic conditions that stress driver situational awareness. A key feature of the simulator is its sensor-fusion–based situational-awareness visualisation, integrating data from virtual cameras, LiDAR, radar, and V2X feeds to recreate how real-world perception systems detect hazards and occlusions. This approach is aligned with state-of-the-art cooperative perception research, which highlights the importance of integrating multi-sensor cues to improve hazard detection and driver awareness [77]. These fused data streams allow the simulator to reproduce realistic edge cases, for example, vehicles hidden in blind zones, or delayed sensor updates that affect handover decisions. Throughout each simulation, the system conducts real-time data logging, capturing driver actions (steering, braking, pedal dynamics), gaze behaviour (fixations, saccades, eyes-on-road time), HMI interactions, reaction times, and physiological indicators where available. Studies consistently show that multimodal logging is essential for evaluating driver readiness, cognitive load, and takeover quality in automated systems [58].

3.4.2.8. Quantitative and qualitative validation criteria matrix

The following table (Table 1) links each validation criterion to its corresponding metric, measurement method, and scientific grounding, enabling traceable, multi-dimensional assessment of human–automation interaction. The approach ensures alignment with state-of-the-art research while integrating proprietary ergonomic and behavioural-simulation tools. The table covers five major evaluation pillars: reaction-time performance, trust dynamics, ergonomics, usability, and perception/arousal. Reaction-time indicators, such as takeover time (TOT), gaze-recovery time, and hands-on-wheel response, are measured in the XR simulator and are grounded in takeover-performance literature [58]. Ergonomic metrics, including RULA scores, joint-load estimations, and posture-risk indicators, following established digital human modelling principles [54]. Usability and UX are assessed using SUS and UEQ, based on ISO 9241 guidance and standard HCI practice for evaluating interactive systems [62]. Finally, perceptual readiness is evaluated through eye-tracking metrics such as pupil-area expansion and gaze fixation patterns, which serve as established arousal indicators in automated driving research [21], [63].

Table 1: Validation Criteria Matrix.

Criterion	Metric	Method
Reaction	Takeover Time (TOT), gaze time, hands-on-wheel time	XR/VR simulator, eye-tracking
Trust	Kalman-based trust score, behavioural trust cues	Real-time behavioural modelling
Ergonomics	RULA, joint torque/load, posture stability	Biophysical digital-human simulation
Usability	SUS, UEQ, clarity & intuitiveness	Post-scenario questionnaires
Perception/Arousal	Pupil area, gaze-shift recovery, fixation	Eye-tracking, pupillometry

3.4.2.9. *Validation output and decision rules*

The final stage of the HMI validation framework involves defining quantitative decision rules and threshold values that determine whether an HMI configuration meets the safety, usability, ergonomic, and trust requirements. These thresholds allow to move from observation to design decisions, ensuring that only interfaces that achieve acceptable performance across all metrics progress to implementation. A primary threshold concerns acceptable takeover performance, which depends on reaction time, gaze recovery, and manoeuvre stability. Prior research shows that takeover times exceeding 5–7 seconds can significantly increase collision risk, especially under high-speed or occluded conditions [58],[60]. Therefore, AviSense.AI uses thresholds such as:

- **TOT ≤ 5 s** under standard conditions
- **TOT ≤ 7 s** under degraded visibility (fog, rain)
- **Hands-on-wheel ≤ 2.0–2.5 s**
- **Eyes-on-road recovery ≤ 1.0–1.5 s**

These values reflect established findings in takeover literature [67] and serve as criteria for accepting AR/HMI alerting strategies. Another decision rule evaluates trust calibration within safe bounds, aiming to avoid both undertrust and overtrust. Studies indicate that miscalibrated trust (e.g., high trust despite poor system performance or persistent distrust of reliable automation) leads to degraded takeover behaviour and increased workload [56], [68]. Acceptable ranges include:

- **Real-time trust score deviations < 20–25%** from normative calibration models
- **No pattern of repeated premature disengagement** (indicator of undertrust)
- **No failure to intervene in hazard scenarios** (indicator of overtrust)

Trust thresholds are computed by comparing subjective trust (Muir, Jian scales) with model-based trust estimates and behavioural cues. Ergonomic thresholds relate to posture safety and musculoskeletal load. AviSense.AI applies established ergonomic standards such as:

- **RULA scores ≤ 3–4** for most interaction tasks (acceptable/moderate risk)
- **Joint-torque values within 30–40% of maximum voluntary capacity**, consistent with musculoskeletal safety studies [49]
- **Head flexion < 20–25°** for HUD/AR-related visual tasks
- **Shoulder elevation < 25°** for most reach interactions

Exceeding these values indicates a potentially unsafe HMI configuration requiring redesign. The framework also defines AR/HMI response-time requirements, ensuring that graphical cues, overlays, and hazard indicators appear with appropriate timing and clarity. Literature on AR-based driving interfaces shows that visual TOR cues should appear ≥ 2.5 s before the minimum safe takeover window, especially in occlusion or low-visibility scenarios. Additional thresholds include:

- **AR cue comprehension time < 1.5 s** (based on gaze and response latency)
- **AR spatial alignment error < 2–3°** in head-tracked AR contexts (supported by AR usability guidelines [67])
- **Information density ≤ 30–40 characters or elements** in the primary FOV to prevent clutter-induced delays [71]

These thresholds indicate that the HMI design adequately supports situational awareness, ergonomic safety, and high-quality takeover performance.

3.4.2.10. XR-based HMI validation framework

XR provides a powerful and flexible framework for validating HMI concepts in automated driving. XR-based validation supports both scientific assessment and training-oriented evaluation, allowing examination of driver performance, situational awareness, reaction time, trust development, and ergonomic safety. A key advantage of XR is the ability to replicate critical scenarios with full environmental control, including fog, rain, night-time glare, occlusions, dense traffic, multi-VRU interactions, and rare hazardous events. Prior research demonstrates that immersive simulation provides high ecological validity while enabling safe exposure to dangerous or unpredictable situations that cannot be directly tested on public roads. This controlled replication is essential for assessing how AR cues, alert modalities, and HMI layouts affect takeover performance, gaze recovery, ergonomic posture, and perception arousal. The XR-based framework also supports dynamic scenario parameterization, allowing to systematically vary traffic density, obstacle behaviour, timing of takeover requests, AR cue placement, and HMI configurations. Immersive driving studies have shown that scenario reproducibility is crucial for comparing interface designs and isolating the influence of individual HMI elements on user performance. By controlling all environmental and event-related variables, the XR simulator enables accurate testing of HMI variants. In addition to its value for evaluation, XR plays a critical role in driver training and skill acquisition. Literature indicates that immersive VR training environments can significantly improve driver hazard anticipation, reduce cognitive load during complex interactions, and increase understanding of automated system behaviours [78].

3.5. Mobile and web applications

3.5.1. Virtual Assistant Interface

CERTH has developed a multimodal cognitive assistant as an embedded ASR-LLM-TTS inference pipeline running entirely on NVIDIA Jetson hardware. Each of the speech pipeline components exposes its functionality over dedicated communication endpoints: the ASR service listens on a dedicated internal port (<IP>:8050) and streams partial and final transcripts to the ROS2 graph, while the TTS service is running on a second port (<IP>:8051), that can be seen in Figure 6, waiting for the synthesized output requests of the LLM node. Internal communication is handled over the Jetson's loopback interface, utilizing isolated namespaces to avoid any traffic collisions. The deployment explicitly separates the conversational interface and the alert interface: the conversation service binds to its own port, handling bidirectional communication and triggering dialogue flows, whereas the alert interface runs in a different container and pushes safety events through another dedicated channel. In this way, alerts can be delivered even if the conversational session is idle or temporarily unavailable.

At runtime, all services communicate over ROS2, but each component is also reachable via defined ports and isolated Docker networks. The containers share one common virtual network for internal routing that enables deterministic message passing and fault isolation between modules. The system can thus easily scale and integrate new front-ends by just assigning new service ports. Based on the requirements of future integration phases, the same architecture can be extended to expose a remote client interface, such as a mobile application, running on a separate port or network endpoint.

3.5.2. Mobile application and cloud platform

The framework designed within the AutoTRUST project will benefit from the vehicle cloud platform developed by ALKE, which offers app and web access and will be customised for the intended use. This platform enables real-time monitoring of vehicle performance, active devices, geotracking (if activated), alerts, diagnostic messages and more.

Figure 14: APP interface for the ALKE cloud platform

3.5.3. Environmental dashboard

UCY is developing a web-based human-machine interface designed to support real-time monitoring of in-cabin environmental conditions. The interface integrates data from air quality, thermal comfort, and related sensors installed on the bus and presents this information to drivers and transport operators to support timely adjustments and operational decisions. Passengers will also be able to view live environmental measurements through a dedicated display, improving awareness of comfort and safety conditions during the trip. The system aims to help maintain a healthy and comfortable cabin environment, with particular attention to the needs of passengers with disabilities.

Figure 15: Integrated digital platform for real-time monitoring and visualisation of in-vehicle environmental data

3.6. Personalization, accessibility and inclusiveness

The AutoTRUST approach to personalization is grounded in the principles of accessibility and inclusiveness, which confirmed the need for interfaces that respect diverse sensory, cognitive, and physical needs while enabling seamless, autonomous mobility for all users. Personalization is not limited to convenience functions; it is a fundamental mechanism through which the vehicle adapts to the user's abilities, preferences, and comfort levels. By integrating both persistent user preferences and real-time adaptation, the system ensures that all passengers, including those with disabilities, sensory sensitivities, or limited digital proficiency, experience a safe and intuitive journey.

3.6.1. Storage of personal preferences

Personal preferences represent the long-term, static traits that we also identified in WP2 and allow the system to configure its baseline behaviour before dynamic adaptation occurs. These preferences may be stored voluntarily by the user through the mobile app, via a profile token (e.g., smartphone, QR code, mobility account), or selected manually during boarding for users who prefer not to store personal data.
Key categories of stored preferences include:

- **Seat preferences**: preferred seating orientation, proximity to doors, seat height or firmness (where adjustable), and anchoring needs for wheelchair users.
- **Climate preferences**: preferred cabin temperature ranges, airflow intensity, and sensitivity-based lighting modes (e.g., low-stimulus settings for sensory-sensitive users).
- **Infotainment preferences**: language choice, audio volume limits, screen-contrast modes, level of detail in visual displays, and opt-in/out for information density.
- **Alert style preferences**: selection of auditory, haptic, or visual notifications depending on sensory abilities, critical for users who cannot rely on one modality alone.

These stored settings allow the system to initialise an environment matched to the user's needs the moment they enter the vehicle. For example, an older adult might receive simplified menus and paced announcements, while a visually impaired or a blind user receives spatialized audio immediately upon boarding

3.6.2. STT/TTS – text display dashboard

The integration of **Speech-to-Text (STT)** and **Text-to-Speech (TTS)** capabilities forms a central component of the AutoTRUST accessibility strategy. In D2.3, multiple user groups, particularly blind and visually impaired participants, older adults, and users with low digital literacy, identified voice interaction as a critical facilitator of independence.

To operationalize this, the system includes an accessible interaction dashboard, a hybrid interface where spoken, visual, and tactile communication converge. This dashboard serves as the central hub for all adaptive interactions between the user and the autonomous vehicle.

3.6.2.1. What is expected from this dashboard?

1. **Multi-modal accessibility**

The dashboard must support simultaneous visual, auditory, and haptic outputs, enabling redundant communication pathways. Visually impaired users can rely on high-quality TTS with clear articulation, while users with hearing impairments receive synchronized text displays or vibration-based alerts.

2. **Adaptable voice characteristics**
   TTS output needs to be adjustable in tone, pitch, speed, and language.
3. **Clear, structured information delivery**
   Information should be broken into predictable segments and avoid simultaneous audio layers, a key requirement from blind and low-vision users in the focus groups. Explanations of maneuvers (e.g., “turning left in 5 seconds”) help reinforce perceived safety.
4. **Real-time STT for users with speech preference**
   STT enables hands-free input for users with mobility impairments or challenges in reaching touch interfaces. The dashboard displays the interpreted command to allow correction or confirmation, preventing system misinterpretations.
5. **Compatibility with screen readers and adaptive visualisation**
   For users relying on assistive technologies, the dashboard provides adaptive visualisation features including high-contrast display modes, large text, reduced visual clutter, and a logical reading order to support screen readers and other assistive tools. In addition, a zoom function allowing magnification of up to 200% is provided to accommodate users with reduced vision or colour blindness. An accessibility mode optimised for touch operation is also required for blind users in order to prevent accidental inputs, following established practices known from modern smartphones.
6. **Explainability and transparency**
   The dashboard communicates *why* certain actions occur (e.g., “Reducing speed for safety,” “Adapting lighting to your sensitivity profile”). This transparency addresses concerns raised by privacy-conscious and first-time AV users.

By linking stored preferences with real-time adaptation, the dashboard becomes an instrument of inclusiveness rather than an obstacle. The data from D2.3 show that personal control, transparency, and alternative interaction modes dramatically improve comfort, perceived safety, and trust especially for users with disabilities and first-time AV users. In this sense, personalization is a central accessibility mechanism ensuring that autonomous mobility remains equitable and intuitive for all.

4. Conclusions

Deliverable D4.2 establishes the conceptual, architectural, and interaction-design foundations of the AutoTRUST VAS and its role in enabling trustworthy, explainable, and personalised human–vehicle interaction. Developed within the framework of WP4 and informed by the user requirements and the multimodal perception capabilities of the Advanced in Cabin Monitoring System, the VAS is conceived as the central cognitive and communicative layer of the AutoTRUST ecosystem. It integrates semantic outputs from in-cabin and external monitoring, contextual reasoning through LLM-based processing, and multimodal user interaction (voice and visual) to provide safe, adaptive, and inclusive interactions.

The deliverable details the initial architecture for the assistant's reasoning pipeline, the integration of speech technologies (ASR/TTS), multimodal perception streams, and preliminary mechanisms for explainability and personalization. These foundations enable the assistant to translate complex sensor outputs into clear and human-understandable feedback, while supporting diverse user needs and ensuring transparency in safety-critical scenarios. Complementarily, the preliminary design activities in Section 3 outline the principles for personalised HMIs, mobile applications, and cross-device interaction flows, ensuring continuity and accessibility throughout the mobility experience.

As this is the first version of the deliverable, the architectural descriptions of both the Virtual Assistant and the HMI ecosystem reflect the current stage of integration and design. The final, fully specified architecture, including finalized module interactions, optimized deployment schemes, complete multimodal interaction pathways, and refined user apps and interface designs, will be provided in the second version of D4.2, following iterative development, integration, and validation activities across WP3, WP4, and WP5.

Co-funded by the European Union

Project funded by Schweizerische Eidgenossenschaft Confédération suisse Confederazione Svizzera Confederaziun svizra Swiss Confederation Federal Department of Economic Affairs, Education and Research EAER State Secretariat for Education, Research and Innovation SERI

KIAT 한국산업기술진흥원 Korea Institute for Advancement of Technology

References

[1] J. Weizenbaum, "ELIZA—a computer program for the study of natural language communication between man and machine," Communications of the ACM, vol. 9, p. 36–45, January 1966.

[2] K. M. Colby, S. Weber and F. D. Hilf, "Artificial paranoia," Artificial intelligence, vol. 2, p. 1–25, 1971.

[3] R. S. Wallace, "The anatomy of ALICE," in Parsing the Turing test: Philosophical and methodological issues in the quest for the thinking computer, Springer, 2007, p. 181–210.

[4] E. Adamopoulou and L. Moussiades, "Chatbots: History, technology, and applications," Machine Learning with applications, vol. 2, p. 100006, 2020.

[5] OpenAI, GPT-4 Technical Report, arXiv, 2023.

[6] G. Team, "Gemini 1.5: Unlocking multimodal understanding across millions of tokens of context," March 2024.

[7] H. Touvron, "Llama: Open and efficient foundation language models," arXiv preprint arXiv:2302.13971, 2023.

[8] M.-B. Group, "AI-powered conversational search," Mercedes-Benz Group, 13 January 2025. [Online]. Available: https://group.mercedes-benz.com/innovations/product-innovation/technology/ai-powered-conversational-search.html. [Accessed 21 November 2025].

[9] V. Group, "ChatGPT is now available in many Volkswagen models," Volkswagen Group, 21 June 2024. [Online]. Available: https://www.volkswagen-group.com/en/articles/chatgpt-is-now-available-in-many-volkswagen-models-18464. [Accessed 21 November 2025].

[10] V. Cars, "Google Gemini is coming to your Volvo with Google built-in," Volvo Cars, 21 June 2024. [Online]. Available: https://www.volvocars.com/en-th/news/technology/google-gemini-is-coming-to-your-volvo-with-google-built-in/. [Accessed 21 November 2025].

[11] S. Media, "Citroën adopts ChatGPT to enhance the on-board comfort experience," Stellantis, 25 July 2024. [Online]. Available: https://www.media.stellantis.com/em-en/citroen/press/citroen-adopts-chatgpt-to-enhance-the-on-board-comfort-experience. [Accessed 21 November 2025].

[12] S. Huang, X. Zhao, D. Wei, X. Song and Y. Sun, "Chatbot and fatigued driver: Exploring the use of LLM-based voice assistants for driving fatigue," in Extended Abstracts of the CHI Conference on Human Factors in Computing Systems, 2024.

[13] V. B. Huber, F. L. Ertl, A. Schlaak, H. Simson and I. Alvarez, "Can an AI Voice Assistant reduce Stress for Drivers confronted with warning Signals?," in Adjunct Proceedings of the 17th International Conference on Automotive User Interfaces and Interactive Vehicular Applications, 2025.

[14] J. Liu, F. Wan, J. Zou and J. Zhang, "Exploring factors affecting people's willingness to use a voice-based in-car assistant in electric cars: An empirical study," World Electric Vehicle Journal, vol. 14, p. 73, 2023.

[15] D. Sogemeier, F. Naujoks, Y. Forster, J. F. Krems and A. Keinath, "Performance and visual appearance of in-vehicle voice assistants impact user experience: A comparative study between Chinese and German users," Transportation Research Part F: Traffic Psychology and Behaviour, vol. 111, p. 279–295, 2025.

[16] A. Chu and G. Huang, "The Intersection of Voice Assistants and Autonomous Vehicles: A Scoping Review," in Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 2024.

[17] C. Bérubé, M. Nißen, R. Vinay, A. Geiger, T. Budig, A. Bhandari, C. R. P. Benito, N. Ibarcena, O. Pistolese, P. Li and others, "Proactive behavior in voice assistants: A systematic review and conceptual model," Computers in Human Behavior Reports, vol. 14, p. 100411, 2024.

[18] A. Bokolo Jnr, "Examining the Use of Intelligent Conversational Voice-Assistants for Improved Mobility Behavior of Older Adults in Smart Cities," International Journal of Human–Computer Interaction, p. 1–22, April 2024.

[19] R. Sangrar, D. J. Murphy and M. M. Porter, "Inclusive autonomous shuttles as public transportation options for older people: A qualitative study of user and service provider perspectives," Journal of Transport & Health, vol. 44, p. 102080, 2025.

[20] J. Wei, X. Wang, D. Schuurmans, M. Bosma, F. Xia, E. Chi, Q. V. Le, D. Zhou and others, "Chain-of-thought prompting elicits reasoning in large language models," Advances in neural information processing systems, vol. 35, p. 24824–24837, 2022.

[21] Q. Wang, H. Chen, J. Gong, X. Zhao and Z. Li, "Studying driver's perception arousal and takeover performance in autonomous driving," Sustainability, vol. 15, p. 445, 2022.

[22] A. Madaan, N. Tandon, P. Gupta, S. Hallinan, L. Gao, S. Wiegreffe, U. Alon, N. Dziri, S. Prabhumoye, Y. Yang and others, "Self-refine: Iterative refinement with self-feedback, 2023," URL https://arxiv. org/abs/2303.17651, 2023.

[23] N. Shinn, F. Cassano, A. Gopinath, K. Narasimhan and S. Yao, "Reflexion: Language agents with verbal reinforcement learning," Advances in Neural Information Processing Systems, vol. 36, p. 8634–8652, 2023.

[24] P. Lewis, E. Perez, A. Piktus, F. Petroni, V. Karpukhin, N. Goyal, H. Küttler, M. Lewis, W.-t. Yih, T. Rocktäschel and others, "Retrieval-augmented generation for knowledge-intensive nlp tasks," Advances in neural information processing systems, vol. 33, p. 9459–9474, 2020.

[25] G. Izacard and E. Grave, "Leveraging passage retrieval with generative models for open domain question answering," in Proceedings of the 16th conference of the european chapter of the association for computational linguistics: main volume, 2021.

[26] H. Zhao, H. Chen, F. Yang, N. Liu, H. Deng, H. Cai, S. Wang, D. Yin and M. Du, "Explainability for large language models: A survey," ACM Transactions on Intelligent Systems and Technology, vol. 15, p. 1–38, 2024.

[27] H. Luo and L. Specia, "From understanding to utilization: A survey on explainability for large language models," arXiv preprint arXiv:2401.12874, 2024.

[28] Y. Jiang, Y. Wang, C. Wu, W. Zhong, X. Zeng, J. Gao, L. Li, X. Jiang, L. Shang, R. Tang and others, "Learning to edit: Aligning llms with knowledge editing," arXiv preprint arXiv:2402.11905, 2024.

[29] L. Monteiro Paes, D. Wei, H. J. Do, H. Strobelt, R. Luss, A. Dhurandhar, M. Nagireddy, K. Natesan Ramamurthy, P. Sattigeri, W. Geyer and S. Ghosh, "Multi-Level Explanations for Generative Language Models," in Proceedings of the 63rd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), 2025.

[30] J. Urakami and K. Seaborn, "Nonverbal cues in human–robot interaction: A communication studies perspective," ACM Transactions on Human-Robot Interaction, vol. 12, p. 1–21, 2023.

[31] P. Wolfert, N. Robinson and T. Belpaeme, "A review of evaluation practices of gesture generation in embodied conversational agents," IEEE Transactions on Human-Machine Systems, vol. 52, p. 379–389, 2022.

[32] P. Y. Cheng, F. Meng, J. Yao and Y. Wang, "Driving With Agents: Investigating the Influences of Anthropomorphism Level and Physicality of Agents on Drivers' Perceived Control, Trust, and Driving Performance," Frontiers in Psychology, vol. 13, p. 883417, June 2022.

[33] NVIDIA, "Audio2Face-3D: Audio-driven Realistic Facial Animation For Digital Avatars," 2024.

[34] W. H. Organization, "Information for Researchers Concerning Informed Decision Mak-ing: What is an Informed Consent Form?," [Online]. Available: https://www.who.int/docs/default-source/documents/ethics/informed-decision-making.pdf..

[35] E. Commission, "Web Accessibility Study," 2020. [Online]. Available: https://ec.europa.eu/social/main.jsp?catId=1202.

[36] W. H. Organization, "Assistive Technology Factsheet," [Online]. Available: https://www.who.int/health-topics/assistive-technology.https://www.who.int/health-topics/assistive-technology..

[37] E. Commission, "European Disability Strategy (2010-2020)," 2015. [Online]. Available: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=LEGISSUM%3Aem0047..

[38] W. H. Organization, "Standards and operational guidance for ethics review of health-related research with human participants," 2011. [Online]. Available: https://iris.who.int/handle/10665/44783..

[39] EN 301 549 V3.2.1: Accessibility requirements for ICT products and services, 2021.

[40] Smart Mobility for Persons with Reduced Mobility — Functional Approaches, DIN 13278:2022. Deutsches Institut für Normung (DIN), 2022. [Online]. Available: https://webstore.ansi.org/standards/din/DIN132782022

[41] A. Naddeo, N. Cappetti and O. Ippolito, "Dashboard reachability and usability tests: a cheap and effective method for drivers' comfort rating," 2014.

[42] M. Capallera, L. Angelini, Q. Meteier, O. A. Khaled and E. Mugellini, "Human-Vehicle Interaction to Support Driver's Situation Awareness in Automated Vehicles: A Systematic Review," IEEE Transactions on Intelligent Vehicles, vol. 8, pp. 2551-2567, 2023.

[43] N. Fahse, M. Millard, F. Kempter, S. Maier, M. Roller and J. Fehr, "Dynamic human body models in vehicle safety: An overview," GAMM-Mitteilungen, vol. 46, April 2023.

[44] H. Yang, Y. Wang and R. Jia, "Dashboard Layout Effects on Drivers' Searching Performance and Heart Rate: Experimental Investigation and Prediction," Frontiers in Public Health, vol. 10, p. 813859, 2022.

[45] A. Naumann and M. Rötting, "Digital human modeling for design and evaluation of human-machine systems," 2007.

[46] I. Grobelna, D. Mailland and M. Horwat, "Design of Automotive HMI: New Challenges in Enhancing User Experience, Safety, and Security," Applied Sciences, vol. 15, p. 5572, 2025.

[47] L. Morra, F. Lamberti, F. G. Pratticó, S. L. Rosa and P. Montuschi, "Building Trust in Autonomous Vehicles: Role of Virtual Reality Driving Simulators in HMI Design," IEEE Transactions on Vehicular Technology, vol. 68, pp. 9438-9450, 2019.

[48] A. Riegler, A. Riener and C. Holzmann, "A systematic review of virtual reality applications for automated driving: 2009–2020," Frontiers in human dynamics, vol. 3, p. 689856, 2021.

[49] A. Naumann and M. Roetting, "Digital Human Modeling for Design and Evaluation of Human-Machine Systems," MMI interaktiv, January 2007.

[50] M. Beggiato and others, "Use of Immersive Virtual Environments for Designing Vehicle Interfaces in Automated Driving," Human Factors, 2021.

[51] L. McAtamney and E. N. Corlett, "RULA: a survey method for the investigation of work-related upper limb disorders," Applied ergonomics, vol. 24, p. 91–99, 1993.

[52] S. Hignett and L. McAtamney, "Rapid entire body assessment (REBA)," Applied ergonomics, vol. 31, p. 201–205, 2000.

[53] K. G. Davis and W. S. Marras, "Evaluation of the LUBA System for Upper-Body Posture Assessment," Ergonomics, 2000.

[54] D. B. Chaffin, Digital human modeling for vehicle and workplace design, SAE International, 2001.

[55] A. Feldhütter, D. Kroll and K. Bengler, "Wake up and take over! the effect of fatigue on the take-over performance in conditionally automated driving," in 2018 21st international conference on intelligent transportation systems (itsc), 2018.

[56] J. D. Lee and K. A. See, "Trust in automation: Designing for appropriate reliance," Human factors, vol. 46, p. 50–80, 2004.

[57] R. Ma and others, "Driver's Perception and Attention in Takeover Transitions During Automated Driving," Transportation Research Part F: Traffic Psychology and Behaviour, 2022.

[58] S. H. Yoon and Y. G. Ji, "Non-driving-related tasks, workload, and takeover performance in highly automated driving contexts," Transportation Research Part F: Traffic Psychology and Behaviour, vol. 60, pp. 620-631, 2019.

[59] L. Pipkorn, M. Dozza and E. Tivesten, "Driver visual attention before and after take-over requests during automated driving on public roads," Human factors, vol. 66, p. 336–347, 2024.

[60] V. Melcher, S. Rauh, F. Diederichs, H. Widlroither and W. Bauer, "Take-Over Requests for Automated Driving," Procedia Manufacturing, vol. 3, pp. 2867-2873, 2015.

[61] P. Bazilinskyy, S. M. Petermeijer, V. Petrovych, D. Dodou and J. C. F. de Winter, "Take-over requests in highly automated driving: A crowdsourcing survey on auditory, vibrotactile, and visual displays," Transportation Research Part F: Traffic Psychology and Behaviour, vol. 56, pp. 82-98, 2018.

[62] J. Brooke and others, "SUS-A quick and dirty usability scale," Usability evaluation in industry, vol. 189, p. 4–7, 1996.

[63] B. Laugwitz, T. Held and M. Schrepp, "Construction and evaluation of a user experience questionnaire," in Symposium of the Austrian HCI and usability engineering group, 2008.

[64] S. G. Hart and L. E. Staveland, "Development of NASA-TLX (Task Load Index): Results of Empirical and Theoretical Research," in Human Mental Workload, vol. 52, P. A. Hancock and N. Meshkati, Eds., North-Holland, 1988, pp. 139-183.

[65] M. A. Regan, K. L. Young and J. D. Lee, Eds., Driver Distraction and Inattention: Principles, Evaluation, and Mitigation, CRC Press, 2008.

[66] V. Venkatesh, M. G. Morris, G. B. Davis and F. D. Davis, "User acceptance of information technology: Toward a unified view," MIS quarterly, p. 425–478, 2003.

[67] H. Azevedo-Sa, S. K. Jayaraman, C. T. Esterwood, X. J. Yang, L. P. Robert Jr and D. M. Tilbury, "Real-time estimation of drivers' trust in automated driving systems," International Journal of Social Robotics, vol. 13, p. 1911–1927, 2021.

[68] B. M. Muir, "Trust Between Humans and Machines: Systems Reliability, Human Reliability, and User Trust," International Journal of Man-Machine Studies, 1994.

[69] S. P. Marshall, "The index of cognitive activity: Measuring cognitive workload," in Proceedings of the IEEE 7th conference on Human Factors and Power Plants, 2002.

[70] O. Palinko and A. Kun, "Exploring the Influence of Light Conditions on Pupil Diameter-Based Workload Estimation," in Proceedings of AutomotiveUI 2011, 2011.

[71] J. Kuhl and D. Fisher, "Physiological Indicators of Driver Alertness in Simulated Driving," Transportation Research Part F: Traffic Psychology and Behaviour, 2014.

[72] S. Kim and A. K. Dey, "Simulated augmented reality windshield display as a cognitive mapping aid for elder driver navigation," in Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, New York, NY, USA, 2009.

[73] J. L. Gabbard, G. M. Fitch and H. Kim, "Behind the glass: Driver challenges and opportunities for AR automotive applications," Proceedings of the IEEE, vol. 102, p. 124–136, 2014.

[74] C. Berthelon and others, "Augmented Reality Cues for Assistance in Driver–Automation Interaction," Transportation Research Part F: Traffic Psychology and Behaviour, 2019.

[75] A. L. Kun, O. Palinko and G. Raz, "Exploring the Effects of Visual Augmented Feedback on Driving Performance," in Proceedings of AutomotiveUI 2013, 2013.

[76] X. Cao and others, "Effects of HUD Display Location and Information Density on Driver Workload and Performance," Applied Ergonomics, 2019.

[77] S. Vora, A. Rangesh and M. Trivedi, "On the Evaluation of Deep Occlusion Reasoning Systems in Autonomous Driving," IEEE Transactions on Intelligent Vehicles, 2020.

[78] N. Mullen, J. Charlton and A. Devlin, "Simulator Validity in Driver Training and Testing," Accident Analysis & Prevention, 2011.