AI-Powered Conversational Assistance in Augmented Reality for Multi-Step Tasks

Juliana H. Madritsch*
Know Center Research GmbH
Graz University of Technology

Tomislav Duricic †
Know Center Research GmbH
Graz University of Technology

Neven ElSayed‡
Know Center Research GmbH

Simone Kopeinik§
Know Center Research GmbH

Eduardo Veas¶
Graz University of Technology
Know Center Research GmbH



Figure 1: The system proposed in this paper. The user is assisted with operating the Juice Mixer testbed with AR visualisations, in terms of cubes, as well as documentation-based AR user interfaces. The baseline user interface operates without AI assistance. Our approach adds a conversational layer to the documentation-based AR system.

Abstract

Augmented Reality (AR) systems that overlay digital instructions onto the physical world can enhance user performance in complex industrial tasks. Procedural instructions can often be derived from existing approved technical documentation, as reuse reduces authoring effort and ensures compliance with established standards. This paper examines the potential of incorporating a Large Language Model (LLM) to add an intelligent, conversational AR assistant to a static AR manual. We conducted an A/B user study ($N = 36$) evaluating the effects of adding a conversational assistance layer for a hands-on task on a juice mixer laboratory installation using a HoloLens 2. The baseline condition provided a PDF manual requiring manual step navigation, complemented by situated visual anchors. The second variant supplemented this interface with a conversational assistant that could interpret user queries, provide direct verbal guidance, and automatically jump to the relevant page in the manual. We evaluated these interfaces across three distinct scenarios: a linear task, a non-linear task requiring users to jump between pages for troubleshooting, and a task-handoff where users had to identify the current state of a partially completed procedure. Our findings indicate that despite longer *total* task completion times, participants using the AI assistant spent significantly less time actively working in the linear and non-linear scenarios, indicating improved task efficiency beyond system latency. Eye-tracking analysis supports the efficiency gain observation as the conversational interface allowed users to focus their visual attention on actively understanding and learning the procedural task. This study highlights the potential of LLM-powered agents for AR guidance, suggesting that overcoming system latency is the critical barrier to their practical deployment in industrial fields.

Index Terms: Augmented Reality (AR), Large Language Models (LLM), Conversational AI, Task Guidance, User Study, Cognitive Load, Context-Aware Systems, Task Handoff.

1 Introduction

Augmented Reality (AR) holds significant promise for transforming industrial operations, particularly for complex procedural tasks such as assembly, maintenance, and training. By overlaying digital information directly onto the user's view of the physical world, AR systems can provide situated guidance, reduce the need for context switching between a workspace and a separate manual, and ultimately improve performance and reduce errors [1, 2, 3]. The ability to access instructions while keeping the hands free is a key, particularly valuable advantage in industrial environments, as it theoretically frees the user to focus entirely on the manual task at hand. In industrial settings, procedural guidance is typically grounded in official technical documentation. For manufactured machinery, standards such as EN ISO 20607 [4] specify the structure and content of operating manuals, including safety-related information.

*e-mail: jmadritsch@know-center.at
†e-mail: tduricic@know-center.at
‡e-mail: nelsayed@know-center.at
§e-mail: skopeinik@know-center.at
¶e-mail: eveas@tugraz.at

These documents are commonly used for training, operation, and maintenance, and serve as authoritative references for prescribed tasks. Consequently, a guidance system deployed in such contexts must adhere to the content and structure of existing documentation. AR-based guidance systems often aim to externalize procedural knowledge from selected sources by presenting instructions directly in the user's field of view. However, procedural documents are commonly authored for sequential reading and may not easily support non-linear execution, conditional operation of steps, or additional information needs that arise during hands-on work. This can lead to increased interaction overhead when users must navigate document structures while simultaneously performing physical tasks. LLMs offer a paradigm shift from static information displays to dynamic, context-aware dialogue systems [5, 6]. Rather than replacing official documentation, LLMs can be used as an intermediate layer supporting query-based access, interpretation, and contextualisation of existing content. In an AR setting, this raises the question of whether conversational interaction offers measurable benefits over traditional document-based presentation while remaining grounded in the same underlying procedural sources. To answer this question, we conducted a comparative user study with 36 participants. We designed and implemented two AR systems for a hands-on procedural task on a physical laboratory testbed: a baseline system that provides a PDF manual with spatial anchors (*Without AI Assistance*), and a system enhanced with a voice-driven LLM assistant (*With AI Assistance*). We chose a document-based AR baseline, despite the known inefficiencies of mid-air document navigation, to reflect common industrial requirements, including the reuse of existing approved technical documentation, which substantially reduces authoring effort. The AI-enhanced assistant supports user queries, provides procedural guidance, and references the corresponding sections of the documentation, thereby preserving traceability to the source material. In this work, we address the following research questions:

RQ1: How does an LLM-powered conversational AR assistant compare to a baseline AR system in supporting procedural tasks?

RQ2: How does task structure (linear, non-linear, conditional) affect the effectiveness of a conversational AR assistant?

RQ3: How does a conversational AR assistant change users' attentional allocation and interaction behaviour during procedural tasks?

To answer these research questions, we conducted a user study comparing a document-based AR manual with an LLM-powered conversational assistant across three scenarios: a linear content navigation task, a non-linear content navigation task requiring troubleshooting and cross-referencing, and a conditional content navigation task in which users had to first assess the system's state before resuming the procedure. We collected a comprehensive set of objective metrics (task completion time, error rates, eye gaze) and subjective measures (cognitive load, usability, user experience).

Our main contributions are: (i) A direct A/B comparison of a baseline AR manual versus an enriched system using a LLM-powered conversational assistant for a complex, hands-on procedural task. (ii) An evaluation across three distinct task types (linear, non-linear, and conditional) that reflect industrial challenges. (iii) An analysis of *effective task completion time*, which isolates user performance from system latency, offering a more nuanced view of the AI assistant's impact. (iv) Eye-tracking evidence demonstrating how conversational AR interfaces alter users' attentional allocation during task execution. Our results indicate that conversational assistance can reduce effective task completion time and visual search effort, particularly in linear and non-linear tasks.

2 Related Work

This work draws from multiple research fields. It is informed by human factors of procedural guidance with multimodal interaction. It builds on the efficacy demonstrated by overlaying task-relevant cues in AR to support procedural tasks. And on the capacity of LLMs to organize, summarize, and generally converse with documents to generate guidance on the fly [7].

Human factors in multimedia procedural guidance. Effective procedural guidance needs to account for how learners process and retain task-relevant information. Eiriksdottir and Catrambone [8] distinguish three instruction types: procedural steps, principles, and examples, noting that step-by-step instructions optimize immediate performance while principles better support transfer. Cognitive load theory offers further constraints: the split-attention principle [9] suggests learners perform worse when integrating spatially separated information, and Mayer's modality principle [10] indicates that auditory narration with graphics can outperform text-with-graphics by distributing load across channels. These principles motivate the spatial integration of instructions and the use of speech to deliver guidance. Beyond the presentation format, navigating information imposes costs. When users navigate complex documentation, they engage in information foraging [11, 12], optimizing information gain against navigation cost. Conversational interfaces might reduce foraging overhead, but whether benefits outweigh interaction costs remains an open question. The Memory for Goals framework [13] addresses interruptions and non-linear navigation: suspended goals lose activation over time, making resumption error-prone without cues. Interruption timing modulates impact [14], and task-switching amplifies costs [15]. For tasks involving handoffs or conditional branching, this predicts resumption difficulties, yet conversational mitigation remains unexplored.

Procedural instructions in Augmented Reality. The efficacy of Augmented Reality (AR) for supporting complex procedural tasks is well established. AR can overlay instructions within the user's view, potentially reducing split-attention demands [1, 16]. Henderson and Feiner [17] demonstrated AR benefits in the psychomotor phase during physical manipulation in military maintenance tasks, showing faster, more accurate assembly versus LCD-based graphics. Subsequent work confirms AR can improve performance [2, 3, 18, 19], but benefits depend on interaction modality and design. Hands-free operation is important [20], and AR's cognitive load effects remain mixed [3, 21], suggesting specific paradigms require individual evaluation. Bahnsen et al. [22] examined AR cues for task resumption, finding reduced resumption lag with both automatic and manual cues, though these were simple markers rather than conversational interventions. Notably, most prior research focused on linear step-following. The demands of complex content navigation, such as non-linear cross-referencing (e.g., troubleshooting) and task handoff/resumption, remain underexplored in head-worn AR [23]. Real-world deployment prioritizes existing documentation rather than pre-authored spatial content, due to authoring costs [24]. Lee et al.[25] introduced DocuBits, which decompose VR documents into spatially positionable elements to improve usability, though manual arrangement is required. Rebol et al. [26] compared AR instruction against videoconferencing for medical procedures, finding different modalities suit different task phases, a consideration for systems like ours that deliver guidance verbally.

LLM conversational task guidance. The limitations of static AR guidance originate in the interaction overhead and the resource-intensive nature of authoring high-quality instructions [24]. Traditional pipelines rely on predefined 3D assets or scene reconstruction [24, 27, 28, 29]. Instead, advances in vision–language models (VLMs/LLMs) offer new paradigms for industrial guidance [30]. Many emerging systems focus on *generating* guidance. Some extract steps from expert videos [24, 29, 31], while others apply LLMs/VLMs to generate natural-language task guidance with-

out bespoke 3D assets [6, 32, 33, 34]. Advanced systems integrate LLMs with real-time vision to produce stepwise, spatially anchored augmentations in the environment [19, 32, 34]. LLMs have enabled AR assistants capable of dialogue [5, 6, 35]. Duricic et al. [31] demonstrated an LLM-powered VR assistant for industrial tasks using a digital twin of manufacturing equipment, processing expert recordings to generate interactive guidance. GazePointAR [36] combined gaze, gesture, and conversation history to disambiguate queries, yet the authors noted challenges in continuous tracking and handling multiple referents. Several systems explore proactive assistance: Persistent Assistant [37] uses grounded input with multimodal feedback; Sensible Agent [38] adapts assistance based on context; AiGet [39] delivers knowledge during casual activities; ProMemAssist [40] models working memory to time interventions. They report promising results in constrained scenarios, though generalisation to arbitrary procedural tasks remains open. Satori [34] proactively generates spatially-anchored guidance by modelling user state. Unlike Satori, our system is reactive; users explicitly query the assistant and work with existing PDFs rather than generating novel content. We chose this design for broader applicability (no task model required) and to preserve user agency, accepting latency costs when users formulate queries.

Collectively, prior work establishes that AR can benefit procedural tasks under certain conditions [1, 17, 18] and that LLMs enable new forms of guidance [5, 6, 19, 29]. However, the specific trade-offs of reactive conversational interaction, querying an assistant to navigate existing documentation, remain underexplored. PDF-based guidance reflects common practice where spatial authoring is unavailable. We contribute an empirical investigation across sequential, non-linear, and handoff tasks, analysing completion times and attention patterns. We believe this perspective is useful for practitioners who must balance existing manuals with emerging conversational support.

3 System Description

3.1 Juice Mixer testbed

The study was conducted using a laboratory installation that simulates an industrial machine, which we call the Juice Mixer testbed [41], shown in Figure 2. The testbed contains six vessels positioned on a vessel stand. Each vessel lid is equipped with three sensors: temperature, pH, and fill level. Tubes connect the vessels with the pumps through a port in the lid. The pumps are activated using the pump control panel switches. Additionally, the flow rate can be set during operation to pump liquids to a designated result vessel.

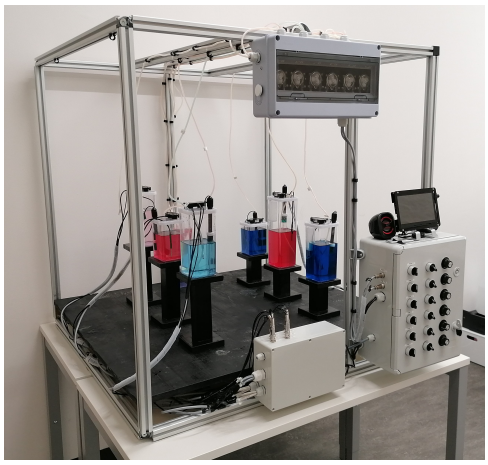


Figure 2: Juice Mixer testbed with a pump control panel and six filled vessels.

3.2 Technical implementation

The application was developed using Unity3D [1] and MRTK 3[2] with HoloLens 2[3] serving as the primary device and streamed to the headset via the Holographic Remoting Player[4].

3.3 AR Interface and Visualisations

To give the users a spatial understanding of the current step, situated AR visualisations (i.e., cubes) have been added for each step. These cubes aid the user in locating parts of the juice mixer or finding the next action positions. The cubes have been placed immersively by an expert variant of our application, where the expert is enabled to choose the number of needed cubes and adjust their position and size, tailored to the necessities of the task. The cubes have been placed for each task and step separately. The positions of the cubes are relative to an attached QR code (dimensions: 14.5 x 14.5 cm) on the juice mixer testbed, which needs to be scanned to activate the task. The cubes are implemented to be gaze sensitive, in terms of changing the colour when looked at (from blue to yellow), and are represented as wireframes, not to occlude the user's sight. An example of such cubes that indicate the positions of the vessels is illustrated in Figure 1.

3.4 Experimental Conditions

3.4.1 Application without AI Assistance

The baseline implementation provides the participants with a PDF manual displayed in an AR-based user interface. To advance through the task, the user can use the buttons *Next Step*, *Current Step*, or *Previous Step*. These buttons result in showing the page of the manual corresponding to the steps, as well as enabling the corresponding AR visualisations aligned with the testbed. The user further has the possibility to manually navigate through the pages and zoom in and out. For switching between chapters, a dropdown menu is provided. These functions are relevant in the non-linear task, where additional information needs to be acquired. The assistant can be closed by pressing the close button on the upper right of the panel, and reopened by pressing a button located in the lower field of the user's view. The assistant panel is shown in Figure 1.

3.4.2 Application with AI Assistance

The application that is enhanced *with AI Assistance*, as presented in Figure 1, includes some changes to the assistant's panel. On the left side of the PDF manual, a chat window is placed for user interaction. To communicate with the assistant, each participant is equipped with a physical button (hand-strapped to the wrist) and a Bluetooth microphone (clipped on clothes). When a user presses and holds the button, their voice is recorded. Once the button is released, the recording is transcribed and shown in the chat window. As soon as the assistant's response is available, it is also displayed in the chat window. In addition to the text output in the chat, users receive a vocal response from the assistant.

3.4.3 LLM Implementation and Configuration

The AI assistant is implemented via the OpenAI Assistant API[5], utilizing an **In-Context Learning (ICL)** strategy. In this approach, the entire knowledge base required by the assistant is provided directly within the system instructions (the prompt context), ensuring that the model's responses are strictly grounded in the provided documentation without requiring external information retrieval.

[1] https://unity.com/de/releases/editor/whats-new/2022.3.33f1

[2] https://learn.microsoft.com/en-us/windows/mixed-reality/mrtk-unity/mrtk3-overview/

[3] https://learn.microsoft.com/de-de/hololens/

[4] https://learn.microsoft.com/de-de/windows/mixed-reality/develop/native/holographic-remoting-player

[5] https://platform.openai.com/assistants/

Model Selection We selected the **GPT-4.1** model. This model was chosen for its superior instruction following capabilities, reliability in producing structured JSON output (essential for tool calling), and a large context window, which was necessary to accommodate the entire manual within the prompt.

Data Input and Preparation The Juice Mixer manual (originally a PDF) was parsed using the PyMuPDF4LLM library and converted into a structured JSONL format. Each entry corresponds to a page, including the page number and markdown content. This entire JSONL structure was injected directly into the assistant's system instructions, serving as its exclusive knowledge base. We used OpenAI's GPT-4o model for integrated speech-to-text (STT) and text-to-speech (TTS) processing.

Structured Output and Reliability To ensure that the assistant's output could be reliably parsed by the Unity application for dynamic control (e.g., automatic page turning, AR highlighting), the response format was strictly set to *json-format*. The temperature was set to 0.01; this low temperature prioritizes deterministic responses, maximizing the reliability of the ICL pipeline over conversational creativity.
The required JSON output schema includes the verbal response (`output`), the source page (`page_in_pdf`), step counts, required AR highlights (`required_boxes`), and the interaction mode (`mode`: "qa" or "guidance").

Assistant Instructions and Prompt Engineering The assistant's behavior is defined by the detailed system prompt containing both the instructions and the knowledge base. Key directives included: (i) *TTS Optimisation:* Instructions emphasized short, clear, and friendly verbal responses. (ii) *Absolute Accuracy and Grounding:* The assistant was strictly confined to the embedded JSONL context. (iii) *Mode Switching:* The assistant was instructed to analyze user intent and switch between a Question Answering ('qa') mode and a Step-by-Step Guidance ('guidance') mode. (iv) *Zero State Handling:* If the user's first prompt implicitly requested a task, the assistant was instructed to immediately begin guidance from Step 1. (v) *Cross-Reference Integration:* The assistant was instructed to automatically include information from cross-referenced sections directly in the output.
This prompt engineering strategy was essential for creating a deterministic assistant capable of driving the AR application logic.

3.5 AI Assistant Performance Characteristics

To contextualize the user study outcomes, we analysed the latency and accuracy of the AI assistant's responses based on the interaction logs during the experiment (N=516 interactions across 18 participants).

Latency: The response time of the assistant (time from the end of the user's utterance to the availability of the AI response) significantly impacted the workflow. The average latency was M = 7.56s (SD = 3.81s). This duration represents the time users spent waiting for the system and is the component subtracted when calculating *Effective Task Completion Time*.

Accuracy: We evaluated the correctness of the assistant's responses based on the defined JSON schema and the knowledge base. A response was considered accurate if it provided the correct information and directed the user to the correct page. Since the assistant relied solely on the provided manual, no hallucinations occurred during the pilot run. However, some responses referenced pages with a consistent offset of minus one. To address this, we added an initial offset check and corrected the page references accordingly. Although the measured accuracy was 66.67%, mainly due to incorrect page numbers, this does not reflect users' perceived accuracy after correction.

4 Experiment

4.1 Study Design

The study, a priori approved by the Ethics Committee of Graz University of Technology, was conducted in a controlled laboratory environment (see Section 3.1), aiming to investigate the use of an LLM-based AR assistant for vocational instruction. Using an A/B study design, we manipulated two independent variables: the presence of an LLM-based assistant (between-subjects) and the task difficulty level (within-subjects). For a comprehensive analysis, the study implements a mixed methods approach, combining qualitative and quantitative measures, with the defined goal to gain insight into the benefits and challenges of using an LLM-based assistant as an extension to AR systems for vocational instruction in industrial environments.
Procedure The study started with an introduction to the experimental details and laboratory equipment. Following this, the participants were asked to sign an informed consent form before being presented with the AR headset and the training task. During the training task, participants were introduced to the user interface and guided on how to proceed through the steps. In line with the experimental condition, the AR assistant was either equipped with a PDF manual (condition *Without AI Assistance*) or a PDF manual plus a conversational agent (condition *With AI Assistance*).

After the training session, participants were instructed to begin the first task as outlined in Section 4.2. While their performance was observed in relation to behavioural observations (see Section 4.4.1), participants completed several questionnaires (see Section 4.4.2) between tasks. These included a Trust in Automation questionnaire (administered before engaging with the study and after each task), the NASA-TLX (completed after each task), as well as the User Experience (UEQ) questionnaires, the System Usability Scale (SUS), and an open-ended post-study questionnaire (completed after finishing the final task).

4.2 Task Design

To evaluate the developed AR guidance systems, we designed three tasks of increasing difficulty. For each task, participants had to follow instructions to mix a juice based on a recipe. They completed the tasks sequentially, from easiest to most difficult.
Standard Procedure. The standard procedure of each task, shown in Figure 3, is subdivided into six steps. The first step involves filling the mixing vessels at the filling station. Then the vessels are placed in the Juice Mixer testbed, followed by attaching the lids, containing the sensors, onto the mixing vessels. After that, the pumps are connected to the vessels by inserting the pump tubes through a port in the lids into the mixing vessels. The fifth step describes the mixing procedure. Here, the pumps are activated by switching the pump activation switches on. The last step is the retrieval of the result vessel from the Juice Mixer testbed.
Task 1: Linear Content Navigation. For this task, the participants are asked to follow a recipe with two mixing vessels. Each step of the recipe follows an uninterrupted standard procedure.
Task 2: Non-Linear Content Navigation. This task was developed to evaluate the usability of the proposed AR guidance system for safety-critical and troubleshooting tasks. The users are asked with cross-references to check for user safety, sensor safety, a malfunctioning pump, and a colour-based quality assessment of the mixed juice in additional chapters. Overall, this task was more difficult for the participants, as they needed to gather the information needed for operation from different chapters.
Task 3: Conditional Content Navigation. For evaluating the usability of the proposed AR systems for handoff tasks, a third recipe was written based on three mixing vessels. The manual for this task is similar to the Linear Content Navigation task. Therefore, a sequential step-by-step guidance was provided, without the need to look for additional information in other chapters.

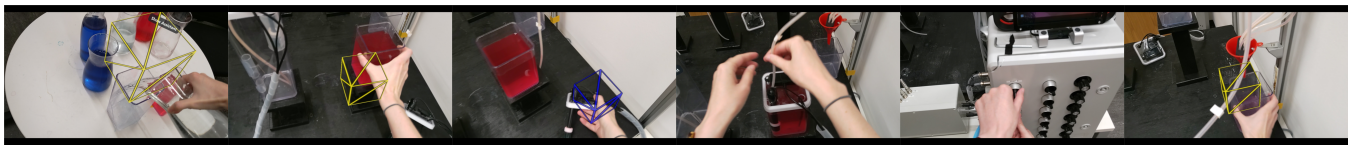


Figure 3: Standard procedure from left to right: Step 1: Filling the vessels for the Juice Mixer, Step 2: Placing the vessels into the machine, Step 3: Attaching lids and sensors, Step 4: Connecting the pump tube to the vessel, Step 5: Activating the pumps of the machine, and Step 6: Retrieving the result vessel.

For this task, the researcher has already prepared parts of the first, second, and third steps. In detail, this means that all three vessels have been filled with the correct liquid. One vessel was correctly placed in the Juice Mixer testbed, and the other two vessels have been switched. The correctly placed vessel was already equipped with the lid, but not connected to the pump. This means the participants had to acknowledge that all three vessels had been filled, switch the positions of the two wrongly placed mixing vessels, and then continue with the standard procedure. This time, all pumps were functioning.

4.3 Participants

A total of 36 participants (14 female, 21 male, 1 non-binary) were recruited through mailing lists and social media postings within the first two weeks of August 2025. To be eligible for participation, participants had to be between 18 and 65 years old. Participants' age ranged from 18 to 55 years (M = 28.94, SD = 8.13). Additionally, the previous usages of AR devices were collected based on a 4-point scale with the items: None = 0, Tried once or twice = 1, Occasional use = 2, Regular use = 3. The AI familiarity was collected based on a 5-point scale with the items: Not at all familiar = 0, Slightly familiar = 1, Moderately familiar = 2, Very familiar = 3, and Extremely familiar = 4. The AR proficiency ranged from "None" to "Regular use" (M = 0.97, SD = 0.91). The AI proficiency ranged from "Slightly familiar" to "Extremely familiar" (M = 2.56, SD = 0.88).
Based on age, gender, familiarity with AR, and familiarity with AI, the participants were split into two independent groups. The participants performing the task in the *Without AI Assistance* condition (7 female, 11 male) were aged from 18 to 55 years (M = 29.28, SD = 10.08), had an AR usage level between "None" and "Regular use" (M = 1.0, SD = 0.91) and an AI familiarity between "Slightly familiar" and "Extremely familiar" (M = 2.67, SD = 0.97). The participants performing the task in the *With AI Assistance* condition (7 female, 10 male, 1 non-binary) were aged from 18 to 41 years (M = 28.61, SD = 5.85), had an AR device usage between "None" and "Regular use" (M = 0.94, SD = 0.94) and an AI familiarity between "Slightly familiar" and "Extremely familiar" (M = 2.44, SD = 0.78).
Each participant was compensated with a 20 Euro gift card for their participation.

4.4 Data Collection

To evaluate user performance, attention distribution, perceived workload, and usability, we collected both objective and subjective data. The following types of data were recorded:

4.4.1 Objective Measures

Task Completion Time: Task completion times were automatically recorded for each task using system logs with millisecond precision. In the *With AI Assistance* condition, we additionally computed an **effective task completion time** by subtracting the time period during which participants were waiting for the AI assistant to respond. This adjustment was made to isolate the participants' active task execution time from system-induced delays.

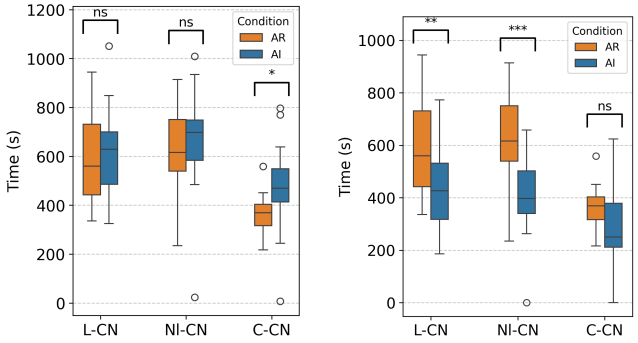
Performance Rate: Performance was calculated based on the number of completed steps. Each task contains six steps. For each step, the total required steps are defined. Performance rate is calculated by dividing the total achieved steps by the total required steps. The performance was recorded during the study by the observer.
Eye Gaze Behaviour: Eye gaze was recorded using the HoloLens 2 device. Gaze metrics are widely used to assess interface usability, attention allocation, and even intent [42]. AOI-based metrics in particular capture search behaviour and resource allocation [43]. On HoloLens 2, there is support for reliable logging and AOI mapping [44], with validated signal quality and calibration procedures [45]; gaze has also been used to infer intent in XR tasks [46]. Areas of Interest were defined for the assistant interface, content display, and other interface regions. Each target hit was recorded with millisecond precision. Based on the target hits, the gaze durations were calculated by calculating the difference between the start and the end time of the gaze event. For each participant and task, we computed the percentage of time spent gazing at the assistant areas of interest relative to the total task duration.

4.4.2 Subjective Measures

NASA-TLX [47]: After each task, participants completed the Raw NASA-TLX questionnaire to assess perceived workload in six categories: mental demand, physical demand, temporal demand, performance, effort, and frustration.
System Usability Scale (SUS) [48, 49]: At the end of the study, participants filled out the System Usability Scale (SUS) to assess the overall usability of the system.
User Experience Questionnaire (UEQ) [50, 51]: Participants also completed the full version of the User Experience Questionnaire (UEQ) at the end of the study, which evaluates six user experience categories: attractiveness, perspicuity, efficiency, dependability, stimulation, and novelty. The scale for the UEQ ranges from -3.0 to 3.0. The semantic scale ranges values above 2.0 as *Excellent*, values between 1.5 and 2.0 as *Good*, values between 0.8 and 1.5 as *Above Average*, values between 0.0 and 0.8 as *Neutral*. All values beneath 0.0 are interpreted as *Negative Experience*.
Post-study open-ended questionnaire: After the study, participants were asked to complete an open-ended questionnaire about confidence, difficulties, and further improvements.
Trust in Automation / AI Questionnaire [52]: Participants filled out a Trust in Automation / AI questionnaire at the beginning of the study and after each task. This questionnaire is outside the scope of this study and will be reported in another work.

4.5 Statistical Data Analysis

For interval- and ratio-scale data, we assessed normality using the Shapiro–Wilk test. When data met the assumption of normality ($p > 0.05$), we used independent samples t-tests to compare the two conditions (*With AI Assistance*, *Without AI Assistance*). In these cases, we reported Cohen's d as the effect size. If the data violated normality assumptions, or if the variable was ordinal (e.g., questionnaire scores), we used the non-parametric Mann–Whitney U test, reporting the rank-biserial correlation as the effect size.



(a) Total task completion times.

(b) Effective task completion times.

Figure 4: Task completion times given in seconds for all three tasks: linear content navigation (L-CN), non-linear content navigation (Nl-CN), and conditional content navigation (C-CN). The significance signs indicate non-significant (ns) and significant results with $p < 0.05$ (*), $p < 0.01$ (**), and with $p < 0.001$ (***).

5 Results

5.1 Objective Measures

5.1.1 Total Task Completion Time

The total task completion time (TCT) was analysed for each task separately. Figure 4a shows the resulting mean TCTs. Since the normality assumption was not violated for all three tasks, independent t-tests were conducted.

Linear Content Navigation: The mean TCT *With AI Assistance* was M = 623.22 s (SD = 187.39 s) while the mean TCT *Without AI Assistance* was M = 587.07 s (SD = 179.41 s). An independent t-test revealed that the difference between them was not significant. The effect size was negligible (Cohen's d = -0.197).

Non-Linear Content Navigation: The mean TCT *With AI Assistance* was M = 666.20 s (SD = 218.77 s) while the mean TCT *Without AI Assistance* was M = 631.15 s (SD = 192.40 s). An independent t-test showed that the difference between the conditions was not significant. The effect size was negligible (Cohen's d = -0.170).

Conditional Content Navigation: The mean TCT *With AI Assistance* was M = 475.39 s (SD = 180.86 s) while the mean TCT *Without AI Assistance* was M = 359.16 s (SD = 83.55 s). An independent t-test (normality assumption was not violated) revealed a significant difference, $t(27.71) = -2.475, p < .05$, with a large effect size (Cohen's d = -0.825).

5.1.2 Effective Task Completion Time

The effective task completion times (eTCT), corrected for assistant latency, were analysed for each task separately. Figure 4b shows the resulting mean eTCTs. Since the data were normal for all three tasks, independent t-tests were performed.

Linear Content Navigation: The mean eTCT *With AI Assistance* was M = 432.05 s (SD = 153.30 s) while the mean eTCT *Without AI Assistance* was M = 587.07 s (SD = 179.41 s). An independent t-test revealed that the difference between them was significant, $t(33.19) = 2.787, p < .01$, with a large effect size (Cohen's d = 0.929).

Non-Linear Content Navigation: The mean eTCT *With AI Assistance* was M = 407.37 s (SD = 154.94 s) while the mean eTCT *Without AI Assistance* was M = 631.15 s (SD = 192.40 s). An independent t-test showed that the difference between the conditions was significant, $t(32.51) = 3.843, p < .001$, with a large effect size(Cohen's $d = 1.281$).

Conditional Content Navigation: The mean eTCT *With AI Assistance* was M = 297.73 s (SD = 143.3 s) while the mean eTCT *Without AI Assistance* was M = 359.16 s (SD = 83.55 s). An independent t-test revealed no significant difference, with a medium effect size (Cohen's d = 0.524).

5.1.3 Task Completion Rate

The performance of the participants was compared between the *With AI Assistance* and the *Without AI Assistance* condition for each task. Since the normality assumption was violated for all tasks, Mann-Whitney U tests were performed.

Linear Content Navigation: The mean completion rate *With AI Assistance* was M = 100% (SD = 0.0%) while the mean completion rate *Without AI Assistance* was M = 98.92% (SD = 2.11%). The difference between both conditions is significant, $U = 198, p < .05$, with a small to medium effect size (rank-biserial r = -0.22).

Non-Linear Content Navigation: The mean completion rate *With AI Assistance* was M = 95.46% (SD = 3.48%) while the mean completion rate *Without AI Assistance* was M = 92.62% (SD = 5.55%). The result shows no significant difference, with a medium effect size (rank-biserial r = -0.30).

Conditional Content Navigation: The mean completion rate *With AI Assistance* was M = 95.99% (SD = 6.82%) while the mean completion rate *Without AI Assistance* was M = 97.84% (SD = 3.88%). The result shows no significant difference, with a small effect size (rank-biserial r = 0.06).

5.1.4 Step Completion Rate

The performance of the participants was compared between the *With AI Assistance* and the *Without AI Assistance* condition for each step of the tasks. Only one step within the tasks shows a significant difference. Since the normality assumption was violated for all steps, a Mann-Whitney U test was performed. The results show a significant difference for the third step of the **Non-Linear Content Navigation** task, $U = 243, p < .001$, with a large effect size (rank-biserial r = -0.5). The effect sizes for all other steps, where no significant difference can be obtained between the *With AI Assistance* and *Without AI Assistant* condition, were small ($|-0.0649| <$ rank-biserial $r < |-0.17|$).

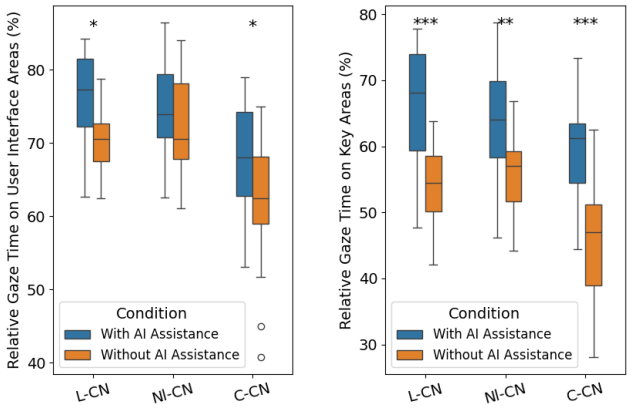
5.1.5 Eye gaze rate on user interface targets

The eye gaze rate was calculated based on all user interface targets, such as the complete assistants containing buttons, a PDF manual, and a chat window, as the AR anchors. Figure 5a shows the distribution as well as significance levels for all tasks and conditions for all user interface targets. Since the normality assumption was not violated, independent t-tests were conducted.

Linear Content Navigation: The mean eye gaze rate on all user interface targets *With AI Assistance* was M = 75.69% (SD = 6.38%) while the mean eye gaze rate on all user interface targets *Without AI Assistance* was M = 70.49% (SD = 5.04%). An independent t-test revealed a significant difference between both conditions, $t(32.26) = 2.71, p < .05$, with a large effect size(Cohen's $d = 0.90$).

Non-Linear Content Navigation: The mean eye gaze rate *With AI Assistance* was M = 74.52% (SD = 6.93%) while the mean eye gaze rate *Without AI Assistance* was M = 72.34% (SD = 6.87%). An independent t-test showed no significant difference between the two conditions, with a small to medium effect size (Cohen's $d = 0.32$).

Conditional Content Navigation: The mean eye gaze rate *With AI Assistance* was M = 67.98% (SD = 7.63%) while the mean eye gaze rate *Without AI Assistance* was M = 61.87% (SD = 9.04%). An independent t-test revealed a significant difference, $t(33.07) = 2.19, p < .05$, with a medium to large effect size (Cohen's $d = 0.73$).



(a) Relative Gaze Time on User Interface Area in percent.

(b) Relative Gaze Time on Key Areas in percent.

Figure 5: Relative Gaze Times for all three tasks: linear content navigation (L-CN), non-linear content navigation (Nl-CN), and conditional content navigation (C-CN). The significance signs indicate significant results with $p < 0.05$ (*), with $p < 0.01$ (**) and with $p < 0.001$ (***).

5.1.6 Eye gaze rate on informative user interface targets

For computing the eye gaze rate on informative user interface targets, only the chat window, the PDF display, and the AR anchors were taken into account. This rate shows the time spent on information retrieval. Figure 5b shows the distribution across tasks and conditions as well as significance levels between the conditions for informative user interface targets. Since the normality assumption was not violated for all tasks, independent t-tests were conducted.

Linear Content Navigation: The mean eye gaze rate on informational user interface targets *With AI Assistance* was M = 65.92% (SD = 9.17%) while the mean eye gaze rate *Without AI Assistance* was M = 54.47% (SD = 5.76%). An independent t-test showed a significant difference, $t(28.60) = 4.49, p < .001$, with a large effect size (Cohen's $d = 1.50$).

Non-Linear Content Navigation: The mean eye gaze rate *With AI Assistance* was M = 64.52% (SD = 8.53%), while the mean eye gaze rate *Without AI Assistance* was M = 55.82% (SD = 5.72%). An independent t-test showed a significant difference, $t(29.73) = 3.59, p < .01$ with a large effect size (Cohen's $d = 1.20$).

Conditional Content Navigation: The mean eye gaze rate *With AI Assistance* was M = 59.98% (SD = 7.9%), while the mean eye gaze rate *Without AI Assistance* was M = 45.57% (SD = 9.35%). An independent t-test revealed a significant difference, $t(33.07) = 4.99, p < .001$ with a large effect size (Cohen's $d = 1.66$).

5.2 Subjective Measures

5.2.1 NASA-TLX sub-scores per task

The values of the NASA-TLX sub-scores range from 0 to 20. The lower the value, the lower the workload. The value for performance was recalculated to be comparable with the other values, as $score_{perf} = 20 - score_{old}$. Since the normality assumption was violated for all tasks, a Mann-Whitney U test was done. Figure 6 visualises the task load for both conditions.

Linear Content Navigation: The mean of the *Mental Demand* sub-score *With AI Assistance* is M = 5.72 (SD = 5.36), while the mean *Without AI Assistance* is M = 6.89 (SD = 5.06). The mean of the *Physical Demand* sub-score *With AI Assistance* is M = 2.94 (SD = 1.43), while the mean *Without AI Assistance* is M = 2.22 (SD = 1.70).

The mean of the *Temporal Demand* sub-score *With AI Assistance* is M = 2.94 (SD = 4.47), while the mean *Without AI Assistance* is M = 2.89 (SD = 3.14). The mean of the *Performance* sub-score *With AI Assistance* is M = 3.5 (SD = 5.14), while the mean *Without AI Assistance* is M = 4.61 (SD = 5.99). The mean of the *Effort* sub-score *With AI Assistance* is M = 5.33 (SD = 4.87), while the mean *Without AI Assistance* is M = 5.00 (SD = 4.37). The mean of the *Frustration* sub-score *With AI Assistance* is M = 6.5 (SD = 5.47), while the mean *Without AI Assistance* is M = 5.56 (SD = 5.18). The difference between the two conditions was not significant. The effect sizes were small ($|0.03|$ <rank-biserial $r < |0.12|$).

Non-Linear Content Navigation: The mean of the *Mental Demand* sub-score *With AI Assistance* is M = 6.11 (SD = 4.73), while the mean *Without AI Assistance* is M = 6.0 (SD = 4.24). The mean of the *Physical Demand* sub-score *With AI Assistance* is M = 2.0 (SD = 1.03), while the mean *Without AI Assistance* is M = 2.67 (SD = 2.81). The mean of the *Temporal Demand* sub-score *With AI Assistance* is M = 2.06 (SD = 1.92), while the mean *Without AI Assistance* is M = 3.11 (SD = 2.99). The mean of the *Performance* sub-score *With AI Assistance* is M = 4.17 (SD = 5.35), while the mean *Without AI Assistance* is M = 4.67 (SD = 5.63). The mean of the *Effort* sub-score *With AI Assistance* is M = 6.17 (SD = 5.61), while the mean *Without AI Assistance* is M = 4.61 (SD = 3.57). The mean of the *Frustration* sub-score *With AI Assistance* is M = 5.72 (SD = 5.17), while the mean *Without AI Assistance* is M = 4.0 (SD = 2.87). No significant difference could be derived. The effect sizes were small ($|0.006|$ <rank-biserial $r < |-0.22|$).

Conditional Content Navigation: The mean of the *Mental Demand* sub-score *With AI Assistance* is M = 4.39 (SD = 3.33), while the mean *Without AI Assistance* is M = 4.33 (SD = 4.24). The mean of the *Physical Demand* sub-score *With AI Assistance* is M = 1.78 (SD = 1.17), while the mean *Without AI Assistance* is M = 2.56 (SD = 2.48). The mean of the *Temporal Demand* sub-score *With AI Assistance* is M = 1.83 (SD = 1.34), while the mean *Without AI Assistance* is M = 2.44 (SD = 1.82). The mean of the *Performance* sub-score *With AI Assistance* is M = 2.11 (SD = 3.56), while the mean *Without AI Assistance* is M = 3.83 (SD = 5.15). The mean of the *Effort* sub-score *With AI Assistance* is M = 4.67 (SD = 4.43), while the mean *Without AI Assistance* is M = 4.22 (SD = 4.15). The mean of the *Frustration* sub-score *With AI Assistance* is M = 5.00 (SD = 5.13), while the mean *Without AI Assistance* is M = 2.89 (SD = 3.25). Between the two conditions, no significant difference was observable.

The effect sizes were small ($|0.09|$ <rank-biserial $r < |-0.28|$).

5.2.2 SUS

The SUS scores can range from 0 to 100; a higher value means a better usability. The mean of the SUS *With AI Assistance* was M = 71.11 (SD = 17.49), while the mean *Without AI Assistance* was M = 70.00 (SD = 19.80). A Mann-Whitney U test showed no significant difference between the two conditions, with a small effect size (rank-biserial r = -0.006).

5.2.3 UEQ

The UEQ was analysed based on the following six sub-scales. The values for the UEQ range between -3.0 and 3.0. A higher value indicates an improved user experience. Figure 7 shows the distributions between conditions. Mann-Whitney U tests have been performed for all subscales. Based on the established benchmarks (see Section 4.4.2), the results indicate a generally Neutral to Negative user experience across both conditions. Only Efficiency and Dependability scored in the positive (Neutral) range (0.0 to 0.8), while Attractiveness, Perspicuity, Stimulation, and Novelty were rated negatively (below 0.0).

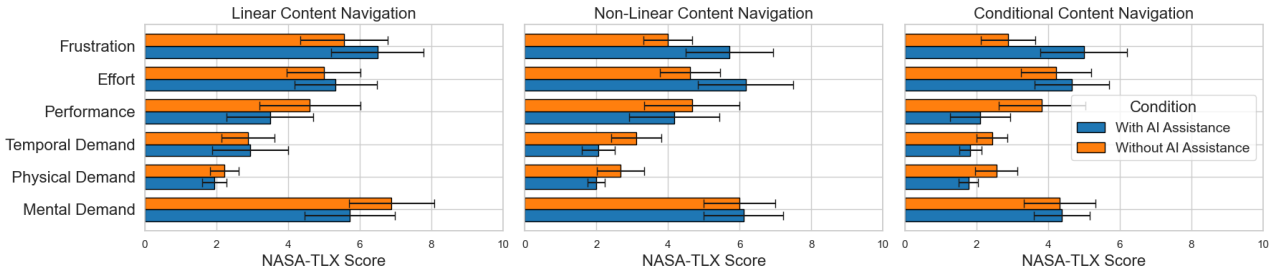


Figure 6: Task load comparison between *With AI Assistance* and *Without AI Assistance*.

Attractiveness: The mean *With AI Assistance* was M = -0.6 (SD = 0.87), while the mean *Without AI Assistance* was M = -0.62 (SD = 0.91). No significant difference was observed with a small effect size (rank-biserial r = 0.04).
Perspicuity: The mean *With AI Assistance* was M = -0.12 (SD = 0.54), while the mean *Without AI Assistance* was M = -0.44 (SD = 0.59). No significant difference was shown with a medium effect size (rank-biserial r = 0.33).
Efficiency: The mean *With AI Assistance* was M = 0.78 (SD = 0.67), while the mean *Without AI Assistance* was M = 0.76 (SD = 0.67). No significant difference was observable with a small effect size (rank-biserial r = -0.03).
Dependability: The mean *With AI Assistance* was M = 0.51 (SD = 0.47), while the mean *Without AI Assistance* was M = 0.53 (SD = 0.53). No significant difference was shown with a small effect size (rank-biserial r = -0.04).
Stimulation: The mean *With AI Assistance* was M = -0.25 (SD = 0.94), while the mean *Without AI Assistance* was M = -0.56 (SD = 0.89). The statistical test revealed no significant difference with a small to medium effect size (rank-biserial r = 0.20).
Novelty: The mean *With AI Assistance* was M = -0.18 (SD = 0.41), while the mean *Without AI Assistance* was M = -0.25 (SD = 0.47). No significant difference was shown with a small effect size (rank-biserial r = 0.07).

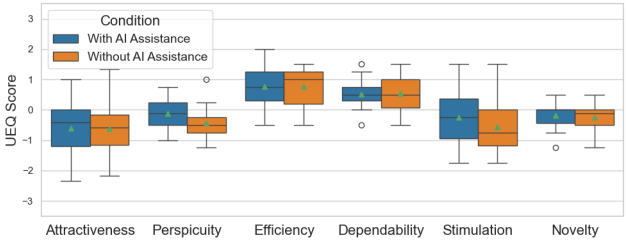


Figure 7: UEQ subscales.

5.2.4 Post-study open-ended questionnaire

With AI Assistance: When asked about the confidence in using the application with AI assistance, three participants stated that they felt extremely confident, 11 were very confident, 3 were moderately confident, and 1 person was slightly confident. Regarding difficulties with the user interface during operation, 7 participants stated that the buttons were either unresponsive or difficult to reach. For two participants, the positioning of the assistant's panel was not useful, as it often occluded their field of view during operation.
Two participants would have preferred fewer visualisations and suggested removing the digital overlay on the Juice Mixer testbed. Two participants would like to see a history of the conversations with the AI assistant, including the images of the PDF manual. When asked if they would use the shown system in a real-world setting, eight participants stated that they would use it in a real-world setting. One participant stated that the system could be useful for furniture assembly. Three participants can imagine using it in a real-world environment. One of these three participants would need a different placement of the assistant panel; the other two would only use it in non-critical environments, where nobody can be endangered. Five people stated that they would not use the system in a real-world setting, as the AI assistant is too slow, or the answers were not accurate enough. One person could not think about a useful application for this system in a real-world setting.
Without AI Assistance: When asked about the confidence in using the application without AI assistance, one person stated extreme confidence, 11 were very confident, 5 were moderately confident, and one participant was slightly confident. Eight participants stated having difficulties with the virtual buttons, and four participants had problems with the floating assistant panel. Three participants had difficulties reading the text in the manual. One participant stated that the digital overlay of the Juice Mixer testbed complicated the vessel placement. Two participants would like to have easier instructions. One participant stated that the usage of a conversational LLM agent could improve the usability. When asked if they would use this system in a real-world environment, fourteen participants stated they would use it. One of these fourteen participants added that it could be helpful for unfamiliar tasks, one would like to see an improved user interface, and another participant would use it if nobody could be endangered during the process. Three participants are not sure if they would like to use the system in a real-world scenario. One participant could not think of a real-world scenario where the system could be used; the other two participants stated that they would use the system with an improved user interface. One participant would not use the system in a real-world application due to the user interface.

6 Discussion

This user study compared an AI-powered conversational assistant with manual content navigation in AR across three distinct task types.
Potential enhancements of using AI-powered AR assistants. (RQ1) The performance results show mixed outcomes when comparing the baseline AR system with the AI-enhanced AR assistant. In terms of Total Task Completion Time (TCT), the AI-enhanced AR system fails to outperform the baseline system, as in the conditional task, users without the AI-enhanced system have been significantly faster than those with. As qualitative feedback indicates a noticeable AI response latency, we additionally evaluated the latency-corrected task completion times (effective TCT). This enables us to theoretically disentangle system latency from task completion times, providing a more nuanced view of the performance of the AI assistant. In terms of effective TCT, the AI-enhanced AR assistant outperformed the baseline AR system in the linear and non-linear tasks, with no significant difference for the conditional task.

This suggests that once information is delivered, the conversational assistant may support more efficient procedural knowledge acquisition, reducing the burden of information foraging [11]. However, the absence of corresponding improvements in subjective usability scores indicates that these efficiency gains were not clearly reflected in overall user experience. Feedback from participants suggests that frustration related to system responsiveness and general AR usability issues may have offset the perceived benefits of the conversational assistant. Overall, the findings indicate that the conversational assistant may support procedural task performance at least as effectively as the baseline AR system under certain conditions, while also highlighting latency as a critical factor influencing real-world usability.

Improved effectiveness of conversational AR assistant for learning tasks. (RQ2) The task structure plays an important role in determining the effectiveness of a conversational AR assistant. Regarding task completion rates, the AI condition outperformed the linear task, while no significant difference could be obtained for the other tasks. This suggests that, especially for training on a new task, the AI-enhanced AR assistant contributes to task completion. Furthermore, this effect was reflected in a higher completion rate for a specific step of the non-linear task, in which the AI condition supported the integration of cross-referenced safety information that was frequently missed in the baseline condition. The lack of significant improvement for the conditional task suggests that when users are familiar with the procedure or only need to confirm the current state, manual content navigation may be sufficient or preferred over engaging in dialogue.

Attentional allocation shift due to conversational AR assistant. (RQ3) Eye-tracking data revealed consistent differences in visual attention between the two conditions. Participants in the AI condition spent a significantly higher percentage of time gazing at informative UI elements (Chat, PDF content, AR anchors) across all tasks. In the baseline condition, users were required to allocate attention to both content consumption and operating the navigation controls (scrolling, menus). In the AI condition, the system handles the navigation, allowing the users to focus their attention on consuming the information provided. While this shift suggests a change in interaction behaviour, subjective feedback indicates that visual focus alone did not necessarily translate into a more positive experience. Participants in both conditions reported occasional occlusion caused by the floating interface as well as user interface issues, which may have constrained effective information consumption regardless of attentional allocation. These results suggest that the AI-enhanced assistant alters how users visually engage with procedural information. The subjective benefits of this shift may depend on the quality and usability of the underlying AR interface.

Design implications for industrial contexts The findings suggest several design implications for deploying AI-enhanced assistance in industrial settings, particularly with respect to scalability, authoring effort, and integration into existing workflows. First, the use of a QR-code-based marker as a reference coordinate system enables a machine-agnostic approach to visualisation. QR codes can be attached to machines without prior modelling or calibration, allowing a rapid deployment across different machines with low installation effort. The use of an immersive expert variant to position generic AR elements provides a low-effort authoring strategy. By relying on simple spatial placeholders rather than detailed 3D models, procedural guidance can be adapted to different machines with minimal configuration. Basing the conversational assistant on existing technical documentation further reduces authoring effort. Showing the original documentation alongside the conversational interface additionally serves as a fallback and as an authoritative reference, which may be relevant for confirmation and compliance in industrial environments. Overall, these design choices support flexible deployment and low authoring overhead.

However, their practical applicability remains dependent on fast AI response times and optimized AR interfaces, as latency and visual interference were found to limit perceived benefits.

7 Limitations

This study has several limitations that should be considered when interpreting the results.

Technical Limitations: The main limitation is the latency of the cloud-based LLM assistant, which directly impacts total task completion time and user experience. Future work could explore on-device LLMs to mitigate latency. Furthermore, the usability issues identified in the HoloLens interface implementation (e.g., button responsiveness, occlusion, text readability) affected both conditions and may have suppressed the potential benefits observed.

Study Design: The study was conducted in a controlled laboratory environment. While the tasks were designed to be ecologically valid (non-linear and handoff scenarios), the results may not fully generalize to real-world industrial settings. Additionally, the repetition of the standard procedure across the three tasks may have introduced learning effects, potentially influencing the results of the final Conditional task.

Scope of the AI Assistant: The AI assistant implemented was a reactive system using In-Context Learning (ICL), relying on user queries and a static manual provided entirely within the prompt context. It did not incorporate visual context awareness (using Vision-Language Models) or proactive guidance strategies (e.g., [34]), which represent promising directions for more advanced assistance.

8 Conclusion and Future Work

This study demonstrates the potential of using conversational AI assistance in AR-guided tasks, allowing more efficient access to procedural information. We show that the benefits of AI assistance are most apparent for users with limited prior familiarity with the procedure. In contrast, as indicated in the conditional content navigation task, no advantage from AI assistance was observed when users were already familiar with operating the test apparatus. This suggests that conversational support is most beneficial in situations where users are exposed to new tasks or are learning a procedure, as well as when they need to integrate various pieces of information. It appears less helpful when users are simply confirming steps they already know in advance. At the same time, the study highlights potential implications for implementing AI-supported AR systems in industrial environments. It identified the need for improvements in AI response times and the quality of answers provided. Future work could explore the use of adaptive AI systems that modify the level of instruction granularity based on user behaviour or task progress. Further, the system could be enhanced by incorporating more images or complementing the AI assistant in displaying images of components or procedures instead of solely directing the user to the correct manual page.

Overall, AI assistance shows promise for supporting procedural tasks, particularly in complex or unfamiliar contexts. However, the absence of significant improvements in users' subjective experience and task completion times underscores the need for more refined AI integration strategies. To fully realize the potential of AI-assisted AR systems in real-world procedural tasks, user-centred design and adaptive interaction mechanisms are essential.

Acknowledgments

The Know-Center is funded within the Austrian COMET Program - Competence Centers for Excellent Technologies - under the auspices of the Austrian Federal Ministry of Transport, Innovation and Technology, the Austrian Federal Ministry of Economy, Family and Youth, and by the State of Styria. COMET is managed by the Austrian Research Promotion Agency FFG.

REFERENCES

[1] A. Tang, C. Owen, F. Biocca, and W. Mou, "Comparative effectiveness of augmented reality in object assembly," in Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, 2003, pp. 73–80. 1, 2, 3

[2] S. Henderson and S. Feiner, "Exploring the benefits of augmented reality documentation for maintenance and repair," IEEE Transactions on Visualization and Computer Graphics, vol. 17, no. 10, pp. 1355–1368, 2010. 1, 2

[3] S. Sudha, V. Keshav Kolla, A. Sanchez, and P. Plapper, "Comparing effectiveness of paper-based and augmented reality instructions for manual assembly and training tasks," in Proceedings of the Conference on Learning Factories (CLF), Graz, Austria, 2021. 1, 2

[4] "Safety of machinery — Instruction handbook — General drafting principles," International Organization for Standardization, Geneva, CH, Standard, Mar. 2019. 1

[5] E. Bozkir, S. Özdel, K. H. C. Lau, M. Wang, H. Gao, and E. Kasneci, "Embedding large language models into extended reality: Opportunities and challenges for inclusion, engagement, and privacy," in Proceedings of the 6th ACM Conference on Conversational User Interfaces, 2024, pp. 1–7. 2, 3

[6] D. Stover and D. Bowman, "Taggar: General-purpose task guidance from natural language in augmented reality using vision-language models," in Proceedings of the 2024 ACM Symposium on Spatial User Interaction, 2024, pp. 1–12. 2, 3

[7] Y. Gao, Y. Xiong, X. Gao, K. Jia, J. Pan, Y. Bi, Y. Dai, J. Sun, H. Wang, and H. Wang, "Retrieval-augmented generation for large language models: A survey," arXiv:2312.10997, vol. 2, no. 1, 2023. 2

[8] E. Eiriksdottir and R. Catrambone, "Procedural instructions, principles, and examples: How to structure instructions for procedural tasks to enhance performance, learning, and transfer," Human factors, vol. 53, no. 6, pp. 749–770, 2011. 2

[9] P. Chandler and J. Sweller, "The split-attention effect as a factor in the design of instruction," British Journal of Educational Psychology, vol. 62, no. 2, pp. 233–246, 1992. 2

[10] R. E. Mayer, Multimedia Learning, 2nd ed. Cambridge University Press, 2009. 2

[11] P. Pirolli and S. Card, "Information foraging," Psychological Review, vol. 106, no. 4, pp. 643–675, 1999. 2, 9

[12] H. Liu, P. Mulholland, D. Song, V. Uren, and S. Rüger, "Applying information foraging theory to understand user interaction with content-based image retrieval," in Proceedings of the Third Symposium on Information Interaction in Context, 2010, pp. 135–144. 2

[13] E. M. Altmann and J. G. Trafton, "Memory for goals: An activation-based model," Cognitive Science, vol. 26, no. 1, pp. 39–83, 2002. 2

[14] P. D. Adamczyk and B. P. Bailey, "If not now, when? the effects of interruption at different moments within task execution," in Proceedings of the SIGCHI conference on Human factors in Computing Systems, 2004, pp. 271–278. 2

[15] E. M. Altmann and J. G. Trafton, "Task interruption: Resumption lag and the role of cues," in Proceedings of the Annual Meeting of the Cognitive Science Society, vol. 26, 2004. 2

[16] R. T. Azuma, "A survey of augmented reality," Presence: Teleoperators & Virtual Environments, vol. 6, no. 4, pp. 355–385, 1997. 2

[17] S. J. Henderson and S. K. Feiner, "Augmented reality in the psychomotor phase of a procedural task," in 10th IEEE International Symposium on Mixed and Augmented Reality, 2011, pp. 191–200. 2, 3

[18] K. M. Baird and W. Barfield, "Evaluating the effectiveness of augmented reality displays for a manual assembly task," Virtual reality, vol. 4, no. 4, pp. 250–259, 1999. 2, 3

[19] A. Y. Zhao, A. Gunturu, E. Y.-L. Do, and R. Suzuki, "Guided reality: Generating visually-enriched ar task guidance with llms and vision models," arXiv:2508.03547, 2025. 2, 3

[20] X. S. Zheng, C. Foucault, P. Matos da Silva, S. Dasari, T. Yang, and S. Goose, "Eye-wearable technology for machine maintenance: Effects of display position and hands-free operation," in Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems, 2015, pp. 2125–2134. 2

[21] F. M. Alessa, M. H. Alhaag, I. M. Al-Harkan, M. Z. Ramadan, and F. M. Alqahtani, "A neurophysiological evaluation of cognitive load during augmented reality interactions in various industrial maintenance and assembly tasks," Sensors, vol. 23, no. 18, p. 7698, 2023. 2

[22] K. L. Bahnsen, L. Tiemann, L. Plabst, and T. Grundgeiger, "Augmented reality cues facilitate task resumption after interruptions in computer-based and physical tasks," in Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems, 2024, pp. 1–16. 2

[23] S. Yoo, S. Reza, H. Tarashiyoun, A. Ajikumar, and M. Moghaddam, "Ai-integrated ar as an intelligent companion for industrial workers: A systematic review," IEEE Access, vol. 12, 2024. 2

[24] A. Stanescu, P. Mohr, D. Schmalstieg, and D. Kalkofen, "Model-free authoring by demonstration of assembly instructions in augmented reality," IEEE Transactions on Visualization and Computer Graphics, vol. 28, no. 11, pp. 3821–3831, 2022. 2

[25] G. Lee, J. Healey, and D. Manocha, "Docubits: Vr document decomposition for procedural task completion," in 2024 IEEE Conference Virtual Reality and 3D User Interfaces (VR), 2024, pp. 309–319. 2

[26] M. Rebol, K. Pietroszek, N. Sikka, C. Ranniger, C. Hood, A. Rutenberg, P. Sasankan, and C. Gütl, "Evaluating augmented reality communication: How can we teach procedural skill in ar?" in Proceedings of the 29th ACM Symposium on Virtual Reality Software and Technology, 2023, pp. 1–11. 2

[27] R. Palmarini, I. F. Del Amo, D. Ariansyah, S. Khan, J. A. Erkoyuncu, and R. Roy, "Fast augmented reality authoring: Fast creation of ar step-by-step procedures for maintenance operations," IEEE Access, vol. 11, pp. 8407–8421, 2023. 2

[28] P. Mohr, B. Kerbl, M. Donoser, D. Schmalstieg, and D. Kalkofen, "Retargeting technical documentation to augmented reality," in Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems, 2015, p. 3337–3346. 2

[29] L. R. Skreinig, P. Mohr, B. Berger, M. Tatzgern, D. Schmalstieg, and D. Kalkofen, "Immersive authoring by demonstration of industrial procedures," in 2024 IEEE International Symposium on Mixed and Augmented Reality (ISMAR), 2024, pp. 1293–1302. 2, 3

[30] W. Fang, Y. Xu, T. Zhang, Y. Wang, and L. Zheng, "Towards large language model for cognitive industrial mixed reality: A survey," Advanced Engineering Informatics, vol. 71, p. 104276, 2026. 2

[31] T. Duricic, P. Müllner, N. Weidinger, N. ElSayed, D. Kowald, and E. Veas, "Ai-powered immersive assistance for interactive task execution in industrial environments," arXiv:2407.09147, 2024. 2, 3

[32] D. Bohus, S. Andrist, N. Saw, A. Paradiso, I. Chakraborty, and M. Rad, "Sigma: An open-source interactive system for mixed-reality task assistance research – extended abstract," in 2024 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW), 2024, pp. 889–890. 3

[33] S. Castelo, J. Rulff, E. McGowan, B. Steers, G. Wu, S. Chen, I. Roman, R. Lopez, E. Brewer, C. Zhao, J. Qian, K. Cho, H. He, Q. Sun, H. Vo, J. Bello, M. Krone, and C. Silva, "Argus: Visualization of ai-assisted task guidance in ar," IEEE Transactions on Visualization and Computer Graphics, vol. 30, no. 1, pp. 1313–1323, 2024. 3

[34] C. Li, G. Wu, G. Y.-Y. Chan, D. G. Turakhia, S. Castelo Quispe, D. Li, L. Welch, C. Silva, and J. Qian, "Satori: Towards proactive ar assistant with belief-desire-intention user modeling," in Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems, 2025, pp. 1–24. 3, 9

[35] X. Wang, T. Kwon, M. Rad, B. Pan, I. Chakraborty, S. Andrist, D. Bohus, A. Feniello, B. Tekin, F. V. Frujeri et al., "Holoassist: an egocentric human interaction dataset for interactive ai assistants in the real world," in Proceedings of the IEEE/CVF International Conference on Computer Vision, 2023, pp. 20 270–20 281. 3

[36] J. Lee, J. Wang, E. Brown, L. Chu, S. S. Rodriguez, and J. E. Froehlich, "Gazepointar: A context-aware multimodal voice assistant for pronoun disambiguation in wearable augmented reality," in Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems, 2024, pp. 1–20. 3

[37] H. Cho, J. Fashimpaur, N. Sendhilnathan, J. Browder, D. Lindlbauer,

T. R. Jonker, and K. Todi, "Persistent assistant: Seamless everyday ai interactions via intent grounding and multimodal feedback," in Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems, 2025, pp. 1–19. 3

[38] G. Lee, M. Xia, N. Numan, X. Qian, D. Li, Y. Chen, A. Kulshrestha, I. Chatterjee, Y. Zhang, and D. Manocha, "Sensible agent: A framework for unobtrusive interaction with proactive ar agents," in Proceedings of the 38th Annual ACM Symposium on User Interface Software and Technology, 2025, pp. 1–22. 3

[39] R. Cai, N. Janaka, H. Kim, Y. Chen, S. Zhao, Y. Huang, and D. Hsu, "Aiget: Transforming everyday moments into hidden knowledge discovery with ai assistance on smart glasses," in Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems, 2025, pp. 1–26. 3

[40] K. Pu, T. Zhang, N. Sendhilnathan, S. Freitag, R. Sodhi, and T. R. Jonker, "Promemassist: Exploring timely proactive assistance through working memory modeling in multi-modal wearable devices," in Proceedings of the 38th Annual ACM Symposium on User Interface Software and Technology, 2025, pp. 1–19. 3

[41] N. ElSayed, O. Prentner, N. Weidinger, and E. Veas, "The juice mixer: A platform for developing and evaluating immersive analytics," in 2025 IEEE Conference on Human Factors in Immersive Analytics (HFIA), 2025, pp. 1–5. 3

[42] X. Yang and I. Krajbich, "A dynamic computational model of gaze and choice in multi-attribute decisions." Psychological Review, vol. 130, no. 1, p. 52, 2023. 5

[43] Y.-B. Zhou, Q. Li, and H.-Z. Liu, "Visual attention and time preference reversals," Judgment and Decision Making, vol. 16, no. 4, pp. 1010—-1038, 2021. 5

[44] S. Kapp, M. Barz, S. Mukhametov, D. Sonntag, and J. Kuhn, "Arett: Augmented reality eye tracking toolkit for head mounted displays," Sensors, vol. 21, no. 6, p. 2234, 2021. 5

[45] S. Aziz and O. Komogortsev, "An assessment of the eye tracking signal quality captured in the hololens 2," in 2022 Symposium on Eye Tracking Research and Applications, 2022, pp. 1–6. 5

[46] A. Seeliger, G. Merz, C. Holz, and S. Feuerriegel, "Exploring the effect of visual cues on eye gaze during ar-guided picking and assembly tasks," in 2021 IEEE International Symposium on Mixed and Augmented Reality Adjunct (ISMAR-Adjunct), 2021, pp. 159–164. 5

[47] S. G. Hart and L. E. Staveland, "Development of nasa-tlx (task load index): Results of empirical and theoretical research," in Advances in psychology. Elsevier, 1988, vol. 52, pp. 139–183. 5

[48] J. Brooke, "Sus-a quick and dirty usability scale," Usability evaluation in industry, vol. 189, pp. 4–7, 1996. 5

[49] ——, "Sus: A retrospective." Journal of usability studies, vol. 8, no. 2, 2013. 5

[50] 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, pp. 63–76. 5

[51] M. Schrepp, A. Hinderks, and J. Thomaschewski, "Construction of a benchmark for the user experience questionnaire (ueq)," International Journal of Interactive Multimedia and Artificial Intelligence, vol. 4, no. 4, pp. 40–44, 2017. 5

[52] S. A. C. Perrig, N. Scharowski, and F. Brühlmann, "Trust issues with trust scales: Examining the psychometric quality of trust measures in the context of ai," in Extended Abstracts of the 2023 CHI Conference on Human Factors in Computing Systems, 2023, pp. 1–7. 5