Comparing AI-Assisted Authoring by Demonstration to Manual Authoring of Augmented Reality Maintenance Instructions

Valentin Tang Knoben
Institute for Anthropomatics and Robotics (IAR)
Karlsruhe Institute of Technology
Karlsruhe, Germany
Institute for Robotics and Autonomous Systems (IRAS)
Karlsruhe University of Applied Sciences
Karlsruhe, Germany
valentin.knoben@kit.edu

Jonas Blattgerste
University of Applied Sciences Emden/Leer
Emden, Germany
jonas.blattgerste@hs-emden-leer.de

Björn Hein
Institute for Robotics and Autonomous Systems (IRAS)
Karlsruhe University of Applied Sciences
Karlsruhe, Germany
Bjoern.Hein@h-ka.de

Christian Wurll
Institute for Robotics and Autonomous Systems (IRAS)
Karlsruhe University of Applied Sciences
Karlsruhe, Germany
Christian.Wurll@h-ka.de

Abstract

Creating Augmented Reality (AR) instructions for industrial guidance remains labor-intensive, while fully automated solutions are often impractical. To bridge this gap, we present VisAARD, an open-source, human-in-the-loop authoring tool that combines vision-language models with authoring by demonstration. From stepwise expert demonstrations, VisAARD uses egocentric video and hand-tracking to generate textual descriptions and propose in-situ 3D guidance elements for later review and refinement. In a within-subject study (N=21), we compared VisAARD against an industry-standard manual authoring baseline similar to Microsoft Dynamics 365 Guides in a robotic-cell maintenance scenario. Compared to the baseline, VisAARD reduced mean authoring time from 36 to under 20 minutes, lowered perceived task load, achieved higher perceived usability, and was preferred by 20 out of 21 participants. Qualitative feedback further suggests that a hybrid workflow combining the efficiency of AI-assisted authoring with targeted manual refinement is perceived as a meaningful direction for AR authoring tools.

CCS Concepts

• **Human-centered computing → Mixed / augmented reality**; *Empirical studies in HCI*; *Interactive systems and tools.*

Keywords

Augmented Reality, Authoring Tools, Authoring by Demonstration, Vision-Language Models, Human-in-the-Loop AI, Industrial Guidance, Maintenance Instructions

MuC '26, Duisburg, Germany

ACM ISBN 979-8-4007-2611-8/26/08
https://doi.org/10.1145/3820253.3820275

ACM Reference Format:
Valentin Tang Knoben, Jonas Blattgerste, Björn Hein, and Christian Wurll. 2026. Comparing AI-Assisted Authoring by Demonstration to Manual Authoring of Augmented Reality Maintenance Instructions. In *Proceedings of Mensch und Computer 2026 (MuC '26), August 30–September 02, 2026, Duisburg, Germany*. ACM, New York, NY, USA, 14 pages. https://doi.org/10.1145/3820253.3820275

1 Introduction

Augmented Reality (AR) has been explored across manufacturing, construction, medicine, and maintenance, with reviews across these domains pointing to clear potential but still limited practical adoption [1, 19, 20, 43]. One particularly promising application of AR is procedural task guidance, where step-by-step instructions are embedded directly in the user's field of view. By presenting information in-situ, AR can support unfamiliar tasks and improve task performance, as it lowers the required cognitive transfer process [13, 14]. Yet despite this potential, adoption in practice remains limited not only due to hardware constraints such as current form factor and computational resources [18], but also by the practical effort required to create AR content [40].

Accordingly, research is increasingly investigating different approaches to simplify the authoring of AR instructions. Recent work conceptualizes this landscape as a still unconsolidated design space in which AR authoring tools differ along numerous dimensions [4]. For the present paper, one particularly useful lens is the level of abstraction offered to the author, ranging from low-level specification of individual instruction elements to higher-level workflows that abstract technical complexity away from the author [22]. Seen through this lens, manual no-code systems, requiring the author to create every single instruction modality for each step by hand, remain the current industry standard as they are broadly applicable, but they are also time-consuming. Converting existing instruction material into AR-based instructions automates the authoring process but depends on the availability of source material in a standardized, high-quality format [28]. Authoring-by-demonstration approaches reduce authoring effort to the physical demonstration of a task and create instructions based on inferred understanding [32]. However, to propose appropriate AR instructions for industrial tasks, they require use-case-specific training data to determine task structure, objects, and context accurately. More recent approaches integrating generative Artificial Intelligence (AI) promise broader generalization by drawing on the pretrained knowledge of general-purpose models rather than task-specific training data [55]. While this is applicable for consumer use cases, this could prove challenging for industrial procedures, which are often highly specific, proprietary, and not represented in publicly available data.

To address this gap, we present VisAARD (Vision Language-supported Authoring of Augmented Reality via Demonstration), an open-source[1], AI-assisted authoring approach in which experts demonstrate a task step by step while a vision-language model interprets each step, drafts corresponding instructions, and proposes their spatial placement. Positioned between fully manual authoring and opaque end-to-end automation, VisAARD keeps the human explicitly in the loop through stepwise review and refinement, preserving author control over content and spatial correctness. This design choice is not only pragmatic given current model limitations but also aligns with human-centric manufacturing and trustworthy-AI perspectives that emphasize human oversight, traceability, and auditability over full automation [26, 37]. In this way, the approach aims to support industrial-type work scenarios without being tied to a single use case or depending on existing instruction material.

The contribution of this specific paper is twofold. First, we contribute VisAARD as an open-source, AI-assisted authoring tool for AR-based procedural maintenance guidance. Second, through a comparative within-subject user study, we investigate the advantages and limitations of AI-assisted authoring by demonstration over manual authoring, used here as an industry-standard baseline closely modeled after Microsoft Dynamics 365 Guides [38], regarding efficiency, perceived task load, usability, and user preference. For this comparison, the evaluated version of VisAARD was deliberately kept close to the core instructional constructs of this baseline so that observed differences primarily reflect the authoring workflow rather than a fundamentally different target instruction format. Through this comparison, we report findings that inform the design of future AR authoring tools and how they can incorporate AI assistance beyond our own specific implementation. We next situate these contributions within prior AR instruction authoring research before detailing the system, user study, and findings.

2 Related Work

While the design space of AR authoring tools is broad and heterogeneous [4], the three categories most relevant to the present paper are AR authoring tools for procedural tasks, AI-assisted AR authoring tools, and AR authoring by demonstration.

2.1 AR Authoring Tools for Procedural Tasks

Starting with procedural AR authoring tools, this category keeps the author in direct control of stepwise AR instructional content but does not rely on any AI assistance. These include Zhu et al.'s ACARS [57], which supports equipment maintenance by combining desktop authoring with on-site AR editing through a mobile user interface and was found to improve efficiency, intuitiveness, and satisfaction through context-aware guidance. Blattgerste et al. [7] proposed an AR authoring tool to support workplace assistance and training by letting instructors create stepwise in-situ AR instructions on a HoloLens using a smartphone controller. They reported good usability and positive reception for both the authoring tool and the resulting AR instructions. In Teach Me How! [21], Funk et al. targeted industrial assembly by deriving interactive in-situ instructions from expert demonstrations using a depth-sensing projection setup. Compared with graphical authoring, it reduced authoring time and workload, while the resulting in-situ instructions were also associated with lower downstream error rates than the video-based assembly instruction and the editor-created in-situ projection instruction. Bégout et al. proposed the AR authoring tool WAAT [11], which targets assembly-line workstation guidance by letting industrial experts configure AR content in a Unity/Vuforia pipeline. In an informal comparison against Microsoft Dynamics 365 Guides, pre-positioned workstation models reduced AR placement effort and marginally improved total authoring time. Lavric et al.'s AR Work Instructions Authoring Tool [33] also targets human-operated assembly lines, but shifts authoring directly into the HoloLens 2 through hand, speech, and gaze input, and found the in-headset workflow substantially faster overall and less complex to use. Finally, TrainAR by Blattgerste et al. [6] supports procedural mobile AR training by giving non-programmers an open-source, visual-scripting-based authoring tool on desktop PCs with deployment to Android and iOS AR devices, and found the tool usable for non-programmers, although usability varied with media competency.

2.2 AI-Assisted AR Authoring Tools

Turning to AI-assisted AR authoring tools, these partially automate generation by delegating parts of authoring to the AI but differ in both the source material they require and the amount of control that remains with the human author. Relying on existing materials, Yamaguchi et al. [53] proposed a tool that turns 2D assembly videos and CAD models into video-annotated AR tutorials that preserve fine-grained tool use and occluded actions. Users preferred it over video tutorials, rating it higher in usability and lower in mental effort. Deng et al. [15] use a multi-agent Vision Language Model (VLM) pipeline to convert user-recorded disassembly videos into AR repair guidance on Meta Quest 3, and an early poster-stage evaluation reported feasible extraction and accurate task descriptions. With lighter-weight author input, TAGGAR by Stover and Bowman [48] uses natural-language instructions and operator photos to ground AR cues to objects on the HoloLens 2 (HL2). It was validated by high end-user task accuracy and rapid operator improvement. CARING-AI by Shi et al. [45] turns spoken or textual task descriptions and contextual screenshots into editable humanoid-avatar instructions and reported good usability as well as fewer mistakes, faster authoring, and lower task load than the baseline method.

The scope broadens further with the ImaginateAR authoring tool by Lee et al. [34], which combines prompts, outdoor scene pre-scans, and in-situ refinement to create AR scenes on an iPhone 13 Pro with LiDAR. Participants preferred hybrid AI-plus-manual control

[1]Source code and system prompts are available at https://github.com/vknoben/VisAARD. The repository at the time of submission is furthermore included in the supplemental materials published in the ACM Digital Library.

in their evaluation. Presenting agentAR, Zhu et al. [56] contributed toward AI-first authoring by generating complete AR applications from natural-language dialogue through autonomous agents on a Quest Pro and a RealSense camera, and its case study and user study reported reduced user effort combined with better planning and tool-utilization performance than the baseline.

Also adjacent is InstructAR by Keelawat and Suzuki [27], which turns text manuals into object highlights, action cues, and step progression, and Guided Reality by Zhao et al. [55], which automatically generates iPad-based AR guidance from user queries and live camera and depth input. Both generate guidance directly for end users rather than supporting reusable authoring workflows and focus on everyday scenarios rather than industrial applications. Closest to our work is ARAIAG by Lin et al. [35], which combines RGB-D assembly demonstrations, no-code engineer input, registration, assembly recognition, and text generated by a Large Language Model (LLM) into industrial AR assembly instructions. In their evaluation, they reported improved text generation after fine-tuning and sufficient registration accuracy for the demonstrated tasks and discussed its preliminary feasibility for industrial authoring.

2.3 AR Authoring by Demonstration

As a subset of AI-assisted AR authoring approaches, authoring-by-demonstration systems reduce authoring effort by capturing a performed task. Previous work here differentiates mainly in terms of hardware used, the interaction concepts, and the amount of post-processing. Whitlock et al.'s AuthAR [52] concurrently captures assembly performance using a HoloLens, a tablet, OptiTrack cameras, and a tracked screwdriver to produce tutorials during task performance. TutorialLens by Kong et al. [31] uses a smartphone, narration, and printed finger markers to turn stepwise device demonstrations into interactive AR tutorials. While AuthAR did not report a formal user evaluation, TutorialLens reported a usable and understandable workflow without requiring AR expertise.

Rather than capturing tutorials directly during task performance, other work reconstructs guidance post hoc from RGB-D demonstrations: AREDA by Bhattacharya and Winer [3] records bench-top assembly with a Kinect and derives part motions for later author refinement, while Stanescu et al. [47] combine HL2 and Azure Kinect to recover geometry, assembly graphs, and interactive tutorials. In both cases, the evaluation focused on technical reconstruction quality and feasibility rather than authoring usability or effectiveness. EditAR by Chidambaram et al. [12] captures a single demonstration as a digital twin with a Quest 2, ZED Mini, and trackers, then shifts the authoring effort into in-headset post-processing to export AR, VR, and video instructions. Compared with traditional single-camera video creation, the recording phase was about five times faster, and EditAR was rated significantly more usable, although it took longer to learn, according to their findings. InstruMentAR by Liu et al. [36] uses HL2, fingertip pressure sensors, and narration to turn embodied demonstrations on digital instruments into arrow-, text-, and image-based tutorials. In a comparative study against manual AR authoring, it reduced authoring time and context switches, achieved good usability, and let users create tutorials faster than the baseline.

Closest to our procedural maintenance task context are Lenssembly by Kosch et al. [32], which uses first-person HoloLens recordings to capture assembly procedures as reusable training instructions, and the immersive authoring approach by Skreinig et al. [46], which uses a Quest 3, a laptop, and optional third-person cameras to demonstrate procedures on virtual machine replicas and semi-automatically segment them into documentation outputs. Lenssembly primarily evaluated the generated playback instructions rather than the authoring process itself, reporting fewer errors and lower workload than paper instructions, while completion times were similar or slower depending on the task. In contrast, Skreinig et al. focused on the authoring tool itself through expert interviews together with usability, workload, and simulator-sickness measures, reporting high usability, low self-reported demand, and positive qualitative feedback on the system's fit for industrial documentation and training workflows.

With TrainAR [6] as a notable open-source exception, the reviewed tools are either not released, have unclear availability to third parties, or depend on technically sophisticated custom setups that limit reproducibility and practical reuse. Moreover, while some prior systems include some evaluation of the authoring process, evidence is often qualitative, small-scale, or intertwined with downstream instruction use and technical validation. As a result, comparative evidence on authoring workflows themselves remains limited, leaving two practical gaps: reusable access to AR authoring tools and stronger comparative evidence on the benefits and tradeoffs of different AR authoring workflows.

3 The VisAARD Authoring Tool

To address these gaps, we present VisAARD, a workflow for authoring AR-based procedural guidance that turns stepwise egocentric demonstrations into draft text and spatially placed 3D instructions by combining demonstration-based capture, VLM-assisted drafting, and explicit human review. Authors remain responsible for demonstrating each work step and reviewing the generated result, while the system assists by interpreting the captured material and proposing corresponding guidance elements. We release the tool as open source to improve reusable access to AR authoring tools, and later compare its authoring workflow against a manual AR authoring baseline. The following section first outlines this workflow and then describes the underlying system design and implementation.

3.1 Authoring Workflow

Figure 1 summarizes VisAARD's authoring workflow from stepwise task demonstration and confirmation to the review, editing, and finalization of the multimodal instructions. The AI-assisted authoring workflow is divided into two key phases: (1) task demonstration and (2) human-in-the-loop review of generated instructions.

3.1.1 Demonstration. For every step of a given task, the author captures an egocentric demonstration of themselves performing the step using the Head-Mounted Display (HMD). The captured egocentric videos serve both as video instructions for their corresponding steps in the final instruction manual and as input to the system to assist with authoring. The author can immediately review the captured video within the stepwise workflow and either

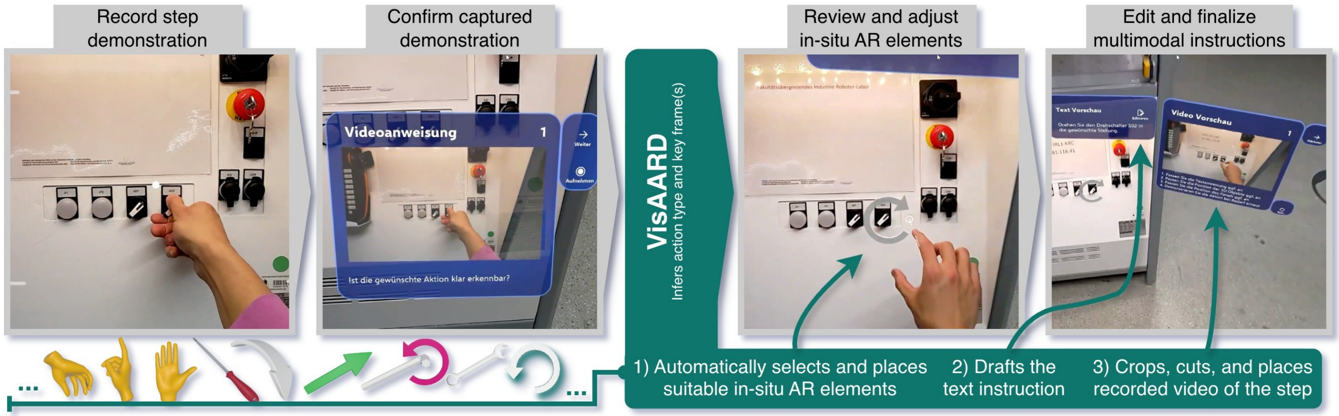

Figure 1: Overview of the two-phase VisAARD authoring workflow. In the demonstration phase, authors record and confirm each task step. VisAARD then generates draft text and spatial AR guidance from the confirmed demonstration while reusing the recorded video for the final instruction. In the review and refinement phase, authors review, edit, and finalize the result.

recapture it or confirm the step. After confirmation, the author continues with the next step while the system processes the captured material in the background to generate accompanying textual and in-situ 3D representations.

3.1.2 Review and Refinement. Having completed the demonstrations, the author enters the review phase. Here, they verify the correctness and placement of the text and 3D instructions proposed by the system. Iterating through each step, the author is presented with their captured video demonstration, a system-generated text instruction, and one or more 3D instructions, positioned automatically by the system. They can modify the text, reposition instruction panels, and manipulate 3D instructions if deemed necessary. Alternatively, the author can choose to re-demonstrate the work step and thereby request a complete regeneration of instructions.

3.2 System Architecture

VisAARD consists of two main components: (1) a HL2-based front-end application and (2) a PC-based back-end application. The HL2 front-end is the interface through which authors capture demonstrations, review generated instructions, and arrange guidance elements. It is implemented in Unity and uses the Mixed Reality Toolkit 3 (MRTK 3) [39] for the design of the user interface and interactions. The back-end processes captured videos, makes requests to the AI APIs, and decides which instructions are suggested to authors. The application is implemented in Python and, during the study, ran on a PC with an Intel Core i9-13900HX CPU, 32 GB RAM, and an Nvidia GeForce RTX 4060. Communication between HL2 and PC builds on a modified server-client setup from Dibene and Dunn [17]. Because this setup does not support control messages from the HL2 server to the PC client, we added another WebSocket channel for HL2-initiated back-end requests. The same setup supports streaming of the HL2 front camera to the PC. The connection between server and client is established before authoring starts. Due to performance constraints, streaming the camera feed reduces the display frame rate from 60 to 30 fps. Spatial registration of holograms relies on the HL2's built-in QR code tracking.

3.3 Demonstration Capture

Video capture is performed locally on the HL2. We set the recording resolution to 1504×846, based on available resolutions on the device, the need for stable 30 fps capture, and initial trials regarding the vision capabilities of current state-of-the-art VLMs. Recording is initiated through hand interaction, whereas stopping is triggered by voice command to avoid visual interference with the captured video. During streaming, the HL2 server notifies the PC client about capture start and end events so that the client can later analyze the relevant temporal segment. After capture ends, the client performs hand detection on the relevant frame window and refines it by removing frames before the first and after the last detected hand appearance, with a small buffer retained on both sides. The adjusted frame range is then communicated back to the server, which trims the recorded video accordingly and presents it to the author for immediate review. In addition to the RGB image stream, the system stores hand-joint data using the built-in hand-tracking capabilities of the HL2. These tracked hand poses are later reused when placing instructions in space.

3.4 Action Understanding

To interpret each captured step, the client samples the incoming frames at a fixed rate of 3 fps. This rate balances the number of frames that can be processed by the VLM with sufficient temporal resolution for short-lived manual actions. Inspired by [50], each sampled frame is automatically numbered with a frame index marker to ensure an unambiguous mapping between input frames and model output. We deliberately adopt a frame-based analysis instead of directly processing video input. This provides explicit control over which frames are analyzed and enables additional preprocessing, such as frame numbering. When given raw video input, VLMs commonly sample it at 1 fps. In our own non-representative preliminary tests, this proved insufficient for understanding short-lived actions spanning only a few seconds, as commonly encountered in maintenance scenarios.

Once the author confirms the recorded step, the numbered frames are forwarded to the VLM. After testing different state-of-the-art VLMs in terms of their visual understanding capabilities, including the latest versions of ChatGPT, Gemini, Claude, Qwen, and Llama at that time, we selected OpenAI's GPT-5.2 model [41] as the analysis component. In the study, we used the OpenAI API model ID `gpt-5.2`. Frames are combined with a system prompt, providing general context, instructions for the VLM, and output constraints. The model is asked to infer two key pieces of information from the demonstration: (1) the action type and (2) the temporal location of the most representative key frame(s). The action vocabulary focuses on common manual handling actions observed in industrial procedures, including pick and place, open and close, press button, rotate switch, use screwdriver, use Allen key, use wrench, fasten by hand, and others. When we experimented with inferring handedness (left/right/both) and direction of rotation (clockwise/-counterclockwise) using the VLM, the results proved unreliable. Therefore, these attributes are determined server-side on the HL2 by analyzing hand motion trajectories and movement distances of tracked hand poses.

3.5 Instruction Proposal and Spatial Placement

Based on the VLM output, VisAARD proposes a textual instruction along with a suitable 3D representation for the analyzed step. The textual instruction is formulated by the model itself, whereas the 3D instruction is chosen according to the detected action type. For actions like pick and place, rotate switch, use screwdriver, use Allen key, use wrench, and press button, VisAARD uses static objects such as arrows, 3D models of tools, or captured hand representations. For all other and undefined actions, the system uses the captured hand motion and renders it as a replay of the observed hand trajectory.

Instruction placement is driven by the hand poses captured during demonstration. Key frames identified by the VLM are first synchronized with the recorded hand-tracking data by selecting the closest corresponding tracking frame. Text and video are combined into a single instruction panel and placed with a vertical offset above the associated hand pose. Static 3D instructions are positioned relative to the same pose using action-specific offsets, for example, by placing an Allen key in the palm or an arrow near the fingertips for pick, place, or rotate. Static hand representations are rendered directly from the recorded pose, whereas hand motion replays use the sequence of hand poses captured within the window of selected key frames.

3.6 Editing and Finalization

After instructions have been automatically proposed, the author reviews the proposed instruction set before it is finalized. Text can be edited via an external physical keyboard, as the built-in holographic keyboard was considered too cumbersome. Static 3D instructions can be adjusted through direct hand-based manipulation as provided through the MRTK 3 interfaces. The combined text and video instruction panel can also be spatially rearranged by the author. 3D instructions based on replayed hand motion are treated as spatially fixed because their representation is directly derived from the recorded hand poses and therefore can only be wrong if spatial registration is inaccurate. If a demonstration is recaptured and resubmitted, the entire instruction analysis and generation pipeline is triggered again. Once the author confirms the reviewed instruction set, the system stores relevant information, including modified textual instructions, panel and object positions, 3D instruction metadata, and recorded video and hand-tracking data, in reloadable formats for future guidance use.

4 User Study

We conducted a comparative mixed-methods user study to evaluate VisAARD's AI-assisted authoring-by-demonstration approach against an industry-standard manual authoring baseline inspired by Microsoft Dynamics 365 Guides [38]. In this section, we outline the manual baseline re-implementation, study use case, design, participant sample, and trial procedure.

4.1 Manual Authoring Tool Re-implementation

To provide a directly comparable baseline, we re-implemented a manual authoring tool within the same application as the AI-assisted mode and reused the same system architecture (see 3.2). The re-implementation was designed in close alignment with Microsoft Dynamics 365 Guides as an industry-oriented reference for manual AR instruction authoring. In this mode, the author manually creates all instruction elements; text is entered via a physical keyboard, video instructions are recorded in the same way as in the AI-assisted workflow, and 3D instructions are assembled from a panel of predefined objects such as arrows, hands, and tools. The author can optionally adjust object color, with white as the default, and then place and orient the selected 3D instruction through hand-based manipulation. The main instruction panel can also be spatially arranged. To proceed to the next step, the author has to write a textual instruction, capture the first-person demonstration, and add at least one 3D object, but the order of these actions is left to the author. This process is repeated for each step until the full workflow is modeled. Although no analysis is required for manual authoring, video streaming to the client is optionally supported, e.g., to allow shared view of the participant's authoring experience. Figure 2 illustrates the described manual authoring workflow.

4.2 Use Case

As the use case for this study, we selected the maintenance of a KUKA Ready-2-Educate robotic cell in our lab. Figure 3 showcases the workplace used for the study. We chose this setting because it offers a realistic industrial-style maintenance scenario while keeping domain knowledge manageable and the study focused on AR authoring. The procedure comprised 11 steps and required participants to open the material box, secure the sideboard using an Allen key, turn on the robotic cell by turning the power switch, open the lateral cell door, insert a cube into the mount inside the cell, insert a pen into the mount inside the cell, close the lateral cell door, acknowledge the door closing by pressing the *Quitt* button, activate the *S1* signal by turning the corresponding switch, activate the *S2* signal by turning the corresponding switch, and close the material box. This covered a broad range of manual handling actions while remaining manageable within one study session.

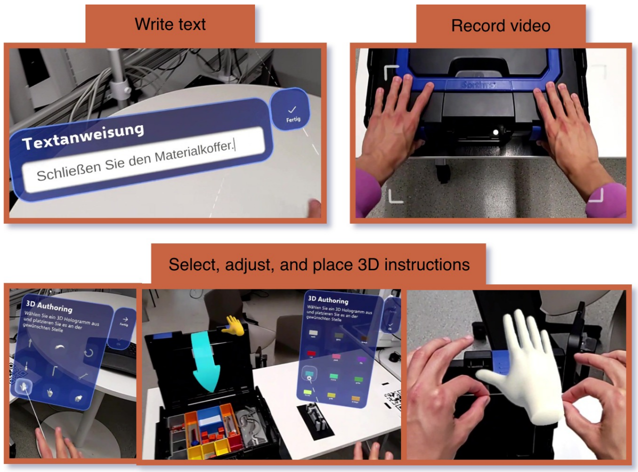

Figure 2: Overview of the manual baseline workflow for one task step: authors can write text, record video, and select, adjust, and place 3D instructions.

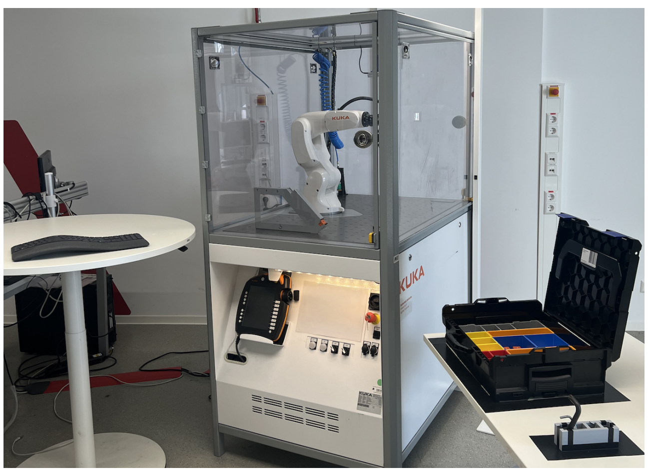

Figure 3: The setup used for the authoring study consisted of a table with a keyboard, a KUKA Ready-2-Educate robotic cell, and a workplace with materials and tools needed.

4.3 Study Design

We employed a within-subject design, counterbalancing condition order across participants to mitigate order effects. Each participant authored the same 11-step robotic-cell maintenance task under both authoring conditions. To keep the comparison as controlled as possible, both conditions were implemented in the same application, used the same hardware and overall system architecture, and relied on the same use case and step sequence. Throughout the study, including the tutorial, the participant's augmented view was streamed to the nearby PC to support guided explanations and observation. Before the main experiment, three pilot tests were conducted to ensure a working prototype and an appropriate overall study duration.

4.3.1 Independent Variable. The examined independent variable was authoring mode, with two levels: *manual authoring* and *AI-assisted authoring*. In the manual condition, participants created all instruction elements by hand (see Section 4.1). In the AI-assisted condition, the system automatically proposes textual instructions, 3D annotations, and spatial placements from human demonstrations (see Section 3.1).

4.3.2 Dependent Variables and Measures. The dependent variables were Task Completion Rate (TCR), Task Completion Time (TCT), number of questions asked, perceived task load, perceived usability, user preference, and qualitative feedback. Task completion and questions asked were tallied manually during the study, with the latter treated as a contextual measure of how often participants sought clarification rather than as a fully standardized performance metric. TCT was logged by the application. Perceived task load and perceived usability were collected after each condition using a non-validated 7-point adaptation of the raw, unweighted NASA-TLX (rTLX) [23, 24] and the System Usability Scale (SUS) [10], respectively. For the rTLX adaptation, we retained the six original NASA-TLX subscales but replaced the original 21-point bipolar response format with the same 7-point Likert response format across both conditions. Responses were linearly rescaled to a 0–100 range and then averaged across the six subscales without applying the original NASA-TLX pairwise weighting procedure. Item wording was adopted from the official NASA-TLX iOS application and translated for the German version of the questionnaire. After completing both conditions, participants indicated their preferred authoring mode and answered open-ended questions about which tool they preferred, why they preferred it, what they particularly liked about AI-assisted authoring, where they saw room for improvement in AI-assisted authoring, what they particularly liked about manual authoring, and where they saw room for improvement in manual authoring. These questionnaire responses were analyzed using a codebook-style thematic analysis [9]. One researcher coded the responses in an initial pass and then conducted a second pass to consolidate codes and build themes.

4.4 Participants

Overall, 21 participants took part in the study. Thirteen were male (61.9%) and eight were female (38.1%). Participants were aged between 25 and 33 ($M = 28.43$, $Mdn = 28.00$, $SD = 2.48$). Regarding prior AR experience, 12 participants (57.1%) reported no prior experience, one participant (4.8%) reported little prior experience, and eight participants (38.1%) reported preexisting experience. With respect to the specific robotic cell used in the study, 11 participants (52.4%) had no prior experience, and 10 participants (47.6%) reported prior experience. The standardized tutorial provided to each participant lasted $M = 10.16\,min$ ($SD = 2.33\,min$).

4.5 Procedure

A full study session lasted between 75 and 140 minutes, including briefing, tutorial, both authoring conditions, and post-task feedback. The study was conducted in accordance with the ethical guidelines of our university. Following prior consultation with our institutional ethics committee, no formal ethics approval was required, as the study posed no physical or psychological risks beyond everyday life and did not involve vulnerable participant groups. All participants were briefed about the study procedure, its purpose, and aspects of data privacy before providing written informed consent. Participants were not monetarily compensated but provided with food and drinks during the session. To protect participant privacy, all data were pseudonymized. Questionnaires referenced participants by assigned ID only, and the mapping table linking IDs to individuals was stored separately from the collected study data.

Each session began with a standardized tutorial scene in which participants familiarized themselves with AR as a technology and its core interaction mechanisms, including hand interaction, text input, video capture, 3D object manipulation, and spatial arrangement of interface elements. Participants then completed both authoring conditions in counterbalanced order, alternating the starting condition across participants to mitigate order effects. At the start of each condition, the experimenter briefly introduced the authoring interface and its functionality. During authoring, the experimenter refrained from intervening and responded only when directly asked by the participant. To convey the task to be authored, participants were provided with third-person reference videos for each step, displayed on the HL2. Third-person perspectives were chosen as they are easy to interpret while avoiding strong bias on how participants capture their own first-person demonstrations. Throughout the study, the participant's augmented view was streamed to a nearby PC to support guided explanations and observations. After completing each condition, participants filled in the rTLX and SUS questionnaires. After both conditions were completed, participants indicated their preferred authoring mode and provided open-ended qualitative feedback.

5 Results

In the following, we report task completion rates, questions asked, authoring time, perceived task load, usability, and user preference, followed by an exploratory analysis of order effects and qualitative feedback. To complement classical inferential statistics, Bayesian paired t-tests were conducted alongside classical paired t-tests to quantify evidential strength, including evidence in favor of null effects [49]. We used the conventional default Cauchy prior for Bayesian t-tests ($r = 0.707$), as proposed by Rouder et al. [42], because no task-specific prior for the expected within-subject effects was available. A sensitivity check with narrower and wider Cauchy priors ($r = 0.5$, 1.0, and $\sqrt{2}$) did not change the qualitative interpretation of the reported Bayesian results. Table 1 summarizes the main quantitative comparisons. Furthermore, following current recommendations for transparent reporting in within-subject designs, visualizations in this section show participant-level data alongside the summary visualizations [51].

5.1 Task Completion & Questions Asked

As summarized in Table 1, all participants completed all 11 steps in both conditions, resulting in a task completion rate of 100% for participants (21/21 in both conditions) as well as for authored steps (231/231 in both conditions). Because task completion showed a ceiling effect, no inferential comparison was applied to this measure.

Participants asked fewer questions in the AI-assisted condition, with a mean of 3.19 questions (SD = 1.66, Mdn = 3.00), than in the manual condition, where they asked 4.38 questions on average (SD = 2.80, Mdn = 5.00). Shapiro-Wilk tests indicated that the paired difference scores were normally distributed (W = 0.927, p = 0.12058), so a paired t-test was used. This difference was not statistically significant, $t(20) = -1.73$, $p = 0.09931$, corresponding to a small within-subject effect ($d_z = -0.38$). The Bayesian paired t-test indicated inconclusive evidence with slight support for the null ($BF_{10} = 0.81$; equivalently, $BF_{01} = 1.24$). In terms of what kind of questions were asked, across both conditions, participants sought clarification on, e.g., general task understanding ($P_{3,5,6,13,14,16,18,21}$), the required level of authoring detail ($P_{12,14,15,17,18}$), or whether physical steps needed to be reversed before moving on to the next step ($P_{5,7,9,15,17}$). Specific to the AI-assisted condition, more frequent inquiries concerned whether it was possible to add or delete 3D instructions ($P_{4,8,10,11,13,17,21}$). In the manual condition, questions focused on the criticality of authoring order ($P_{3,4,14,15,21}$) and the mechanisms of remanipulation ($P_{2,4,8,9,15,18,21}$) or deletion of 3D objects ($P_{1,7,8,16,18,19}$) after initial creation.

5.2 Task Completion Time

As shown in Table 1 and Figure 4, authoring time was substantially lower with AI-assisted authoring, with a mean of 19 min 33 s (SD = 5 min 26 s, Mdn = 18 min 06 s), than with manual authoring, which required 36 min 00 s on average (SD = 10 min 50 s, Mdn = 31 min 20 s). Shapiro-Wilk tests indicated that the paired difference scores were normally distributed (W = 0.966, p = 0.63748), so a paired t-test was used. This difference was significant, $t(20) = -7.17$, p = 0.00000061, corresponding to a large within-subject effect ($d_z = -1.56$). The Bayesian paired t-test indicated extreme evidence for a difference (BF_{10} = 28,390). The time advantage of AI-assisted authoring persisted across all steps of the authoring process, with lower mean times at every step. As shown in the right panel of Figure 4, the between-condition gap was descriptively most pronounced in the first two steps and remained smaller, although variable, across the later steps.

5.3 Perceived Task Load

A similar pattern emerged for perceived task load: rTLX scores were lower in the AI-assisted condition than in the manual condition (see Table 1 and the left panel of Figure 5), with a mean of 15.34 (SD = 9.81, Mdn = 13.89) compared to 26.19 (SD = 13.16, Mdn = 25.00). Shapiro-Wilk tests indicated that the paired difference scores were normally distributed ($W = 0.961$, $p = 0.53672$), so a paired t-test was used. This difference was significant, $t(20) = -4.45$, $p = 0.00024749$, corresponding to a large within-subject effect ($d_z = -0.97$). The Bayesian paired t-test indicated extreme evidence for a difference (BF_{10} = 123.45).

5.4 Usability & Preference

Usability ratings likewise favored AI-assisted authoring (Table 1, right panel of Figure 5), with a mean SUS score of 86.67 (SD = 11.55, Mdn = 90.00) compared to 79.05 for manual authoring (SD = 11.79, Mdn = 75.00). Shapiro-Wilk tests indicated that the paired difference scores deviated from normality ($W = 0.882$, $p = 0.01583$), so a Wilcoxon signed-rank test was used. This difference was significant, $W = 38.00$, $p = 0.02152$, corresponding to a large effect ($r_{rb} = 0.60$).

Table 1: Main quantitative comparisons between AI-assisted and manual authoring. Significant p values are shown in bold.

Outcome	AI-assisted			Manual			Test	Statistic	p	Effect size	BF_{10}
	M	SD	Mdn	M	SD	Mdn					
TCR	100% / 100%			100% / 100%			*(No inferential comparison due to ceiling effect)*				
Questions asked	3.19	1.66	3.00	4.38	2.80	5.00	Paired t	$t(20) = -1.73$	0.099	$d_z = -0.38$	0.81
TCT (min:s)	19:33	5:26	18:06	36:00	10:50	31:20	Paired t	$t(20) = -7.17$	**< .001**	$d_z = -1.56$	28,390
rTLX	15.34	9.81	13.89	26.19	13.16	25.00	Paired t	$t(20) = -4.45$	**< .001**	$d_z = -0.97$	123.45
SUS	86.67	11.55	90.00	79.05	11.79	75.00	Wilcoxon	$W = 38.00$	**0.022**	$r_{rb} = 0.60$	–
Preference	20/21 (95.2%)			1/21 (4.8%)			*(No inferential comparison due to ceiling effect)*				

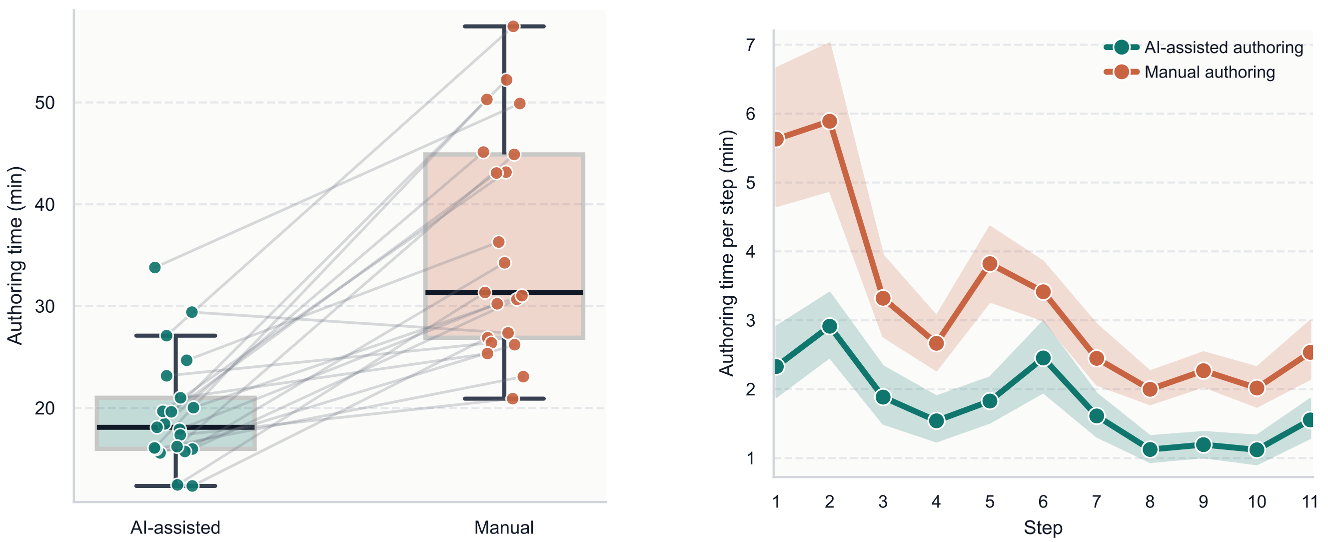

Figure 4: Authoring time results. Left: overall authoring time by condition with participant-level trajectories and condition distributions. Right: mean authoring time per step for AI-assisted and manual authoring with 95% CI (lower is better).

No Bayesian test was added here because normality was violated. Contextualizing the SUS scores using the SUS Analysis Toolkit [5], the AI-assisted score corresponds to "best imaginable usability" on the adjective scale, whereas the manual score corresponds to "good usability". Both scores are above the average SUS score of 68 often identified in meta-analyses, but the AI-assisted condition also exceeds the benchmark of 80, which is often used as an industry standard.

At the end of the study, 20 of 21 participants (95.2%) stated that they preferred AI-assisted authoring, whereas one participant preferred manual authoring. Given this extreme observed proportion, preference is reported descriptively, and no additional inferential or Bayesian test was applied.

5.5 Order Effects

Order effects were examined by comparing participants who started with AI-assisted authoring ($n = 11$) and participants who started with manual authoring ($n = 10$). Descriptively, the clearest pattern appeared for TCT: the AI-assisted advantage was larger for participants who started with manual authoring ($M_{diff} = -1480.52\,s$, $SD = 457.77$) than for those who started with AI-assisted authoring ($M_{diff} = -538.68\,s$, $SD = 378.01$).

For rTLX, both start-order groups showed lower scores for AI-assisted authoring than for manual authoring, and the size of this difference was visually similar across orders (AI-assisted first: $M_{diff} = -13.13$, $SD = 13.03$; manual first: $M_{diff} = -8.33$, $SD = 8.69$). For SUS, the most pronounced pattern concerned the manual condition: participants who started with AI-assisted authoring rated manual authoring lower ($M = 71.36$, $SD = 8.76$) than participants who started with manual authoring ($M = 87.50$, $SD = 8.50$). For questions asked, participants who started with manual authoring showed a stronger reduction when switching to AI-assisted authoring ($M_{diff} = -3.40$, $SD = 2.91$) than participants who started with AI-assisted authoring, for whom the pattern was comparatively small and in the opposite direction ($M_{diff} = 0.82$, $SD = 1.72$). These patterns are plotted in Figure 6 but are exploratory and reported to contextualize possible learning and contrast effects in the discussion rather than show confirmatory findings.

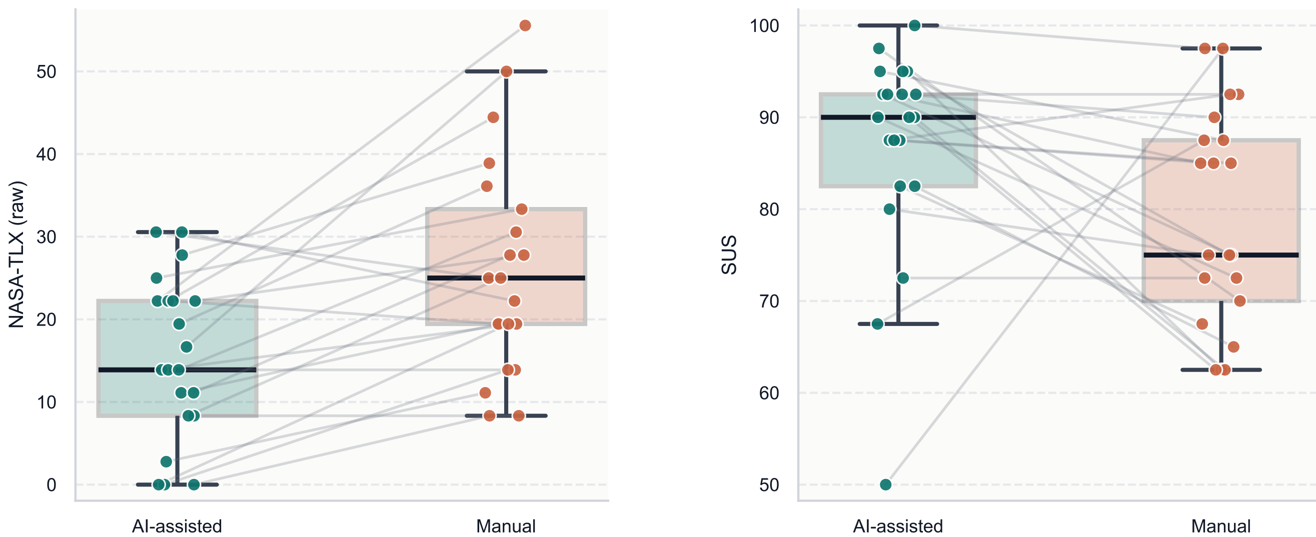

Figure 5: Subjective ratings by condition. Left: perceived task load with participant-level trajectories and condition distributions (lower is better). Right: perceived usability with participant-level trajectories and condition distributions (higher is better).

5.6 Qualitative Feedback

The open-ended, post-study feedback revealed several recurring themes. Direct quotes are translated into English, except for $P_{5,12}$, whose feedback was already provided in English. Participants primarily grounded their preference for AI-assisted authoring (see 5.4) in efficiency gains, both in speed and workload (P_{10}, P_{11}, P_{13}–P_{21}). Easier use and provided points of reference in the form of suggested instructions were additional specified reasons (P_{9-11}). P_9 summarized their preference for AI-assisted authoring as follows: "Just making changes is much more intuitive. You already get a sense of the content and don't have to worry about the wording anymore. The 3D objects were also already included in the automated version, which made things much easier". P_{12} was the only person preferring manual authoring due to its greater freedom in authoring.

Feedback on the strengths and weaknesses of AI-assisted authoring provided further insights that largely coincided with the stated preferences. The main positive contributing factors were increased efficiency ($P_{1-7,11,12,15,20}$) and animated 3D instructions ($P_{2,6,8,9,13,14,16,18,19}$). Having suggested instructions as a starting point for authoring was described as very beneficial ($P_{5,7,9,10,17,21}$). Key points of criticism were imprecise placement of 3D instructions and inconsistency in generated text across steps ($P_{6,9,11,12,14,18-20}$). P_{14} stated that "There was some imprecise phrasing and incorrect placement of 3D objects". Lack of authoring options, such as the possibility to add or delete 3D objects to the system-proposed instructions, was criticized by $P_{4,8,21}$. $P_{1-3,7,10}$ asked for further features such as automatic spelling checks and multiple distinct instruction suggestions.

On the other hand, what sticks out positively for manual authoring is that participants felt less constrained in their way of authoring ($P_{2,4,5,8,11,12,14-16,19-21}$) and were able to match their expectations in instructional quality ($P_{8,12,15,20}$). P_{12} describes the benefits of manual authoring as follows: "Having the agency of being as descriptive as possible. For example, I can be as descriptive as I can be with the text, then create a video demonstration matching closely to the text instruction, followed by enhancing the clarity with the symbols". $P_{1,3,6,7,9}$ did not specify strengths of the manual approach or restate their preference for AI-assisted authoring. Usability, mostly the placement and manipulation of 3D objects ($P_{10,14,16,17,19,20}$) and frustrating UI interactions ($P_{3,4,6,13,15,17}$), were the primary points of dissatisfaction. Inefficiency of authoring was also criticized explicitly ($P_{11,12,15}$). Additional features requested in the course of this were quick modifications such as orientation flips and copying of 3D and text instructions ($P_{13,15,21}$), the creation of animated instructions ($P_{2,5,8,12,18}$), or a checklist of what is left to author for each step (P_1).

6 Discussion

In the following, we contextualize and interpret our findings beyond the reported results. We begin by more closely examining the meaning behind study data, order effects, and found outliers, followed by a reflection on the nature of authored instructions and limitations of our study. Subsequently, we discuss implications for AR authoring tools, covering a possible hybrid authoring approach, the evolving role of the human in the AR authoring process, and future research directions in that regard.

6.1 Interpretation of Empirical Findings

Across all measured outcomes apart from the number of questions asked, the results exhibit a significant advantage of AI-assisted over manual authoring. Most notably, AI-assisted authoring reduced active authoring time by 987 s on average, corresponding to approximately 16 min 27 s saved per task. This is not only an empirically significant result but would also have a substantial impact in practice. If authors are required to author instructions for multiple or

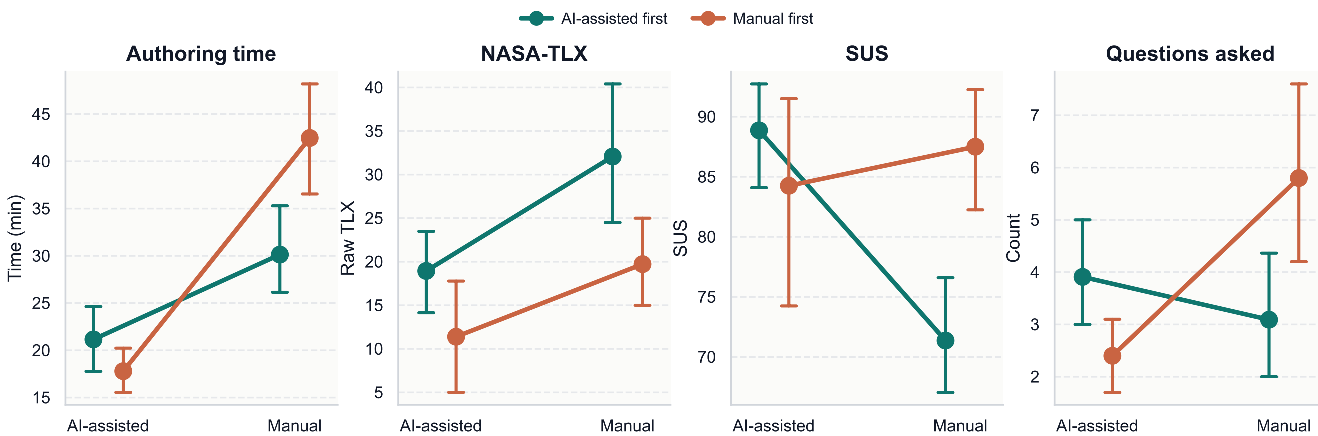

Figure 6: Exploratory order effects by starting condition. Lines show condition means with 95% CI for participants who started with AI-assisted authoring and participants who started with manual authoring. Lower values are better for authoring time, rTLX, and questions asked; higher values are better for SUS.

simply more complex tasks consisting of a larger number of steps, effects on authoring time and perceived task load could potentially be even more pronounced. Although not examined in this user study, keeping the overall authoring time as short as possible could also positively impact instruction quality. The influence of fatigue on output quality, e.g., in manufacturing, has been shown [54]. The same appears reasonable for authoring instructions.

The perceived usability and task load results complement these findings and suggest that the observed efficiency gains meaningfully contributed to a more positive user experience. This is reflected in the shift of mean SUS score from *good* (79.05) to *best imaginable* (86.67) [5], combined with a substantially lower perceived task load in the AI-assisted condition ($d_z = -0.97$, $BF_{10} = 123.45$), indicating that participants not only completed the authoring task faster but also experienced the process as less demanding and more user-friendly. In that regard, qualitative feedback points to having a basis for authoring and the animated 3D instructions as key contributors.

The one outcome without a clear difference was the number of questions asked. This may partly reflect that each condition introduced its own category of uncertainty rather than one being uniformly clearer than the other. While both conditions shared common patterns of seeking clarification, such as general task understanding or desired level of instruction detail, the nature of questions appeared to shift between conditions. In the manual condition, questions centered on how to use existing features, such as manipulation and deletion of 3D objects. In contrast, in the AI-assisted condition, questions took on a more suggestive character, pointing to functionality participants expected but found absent, such as the ability to add or delete 3D instructions during review.

6.1.1 Order and Learning Effects. The clearest exploratory order-related pattern appeared for authoring time. Participants who started with manual authoring exhibited substantially higher overall authoring times than those who started with AI-assisted authoring, which is consistent with a learning effect benefiting the second condition. Notably, this effect is unlikely to stem from increased familiarity with the task itself, as watching and understanding the demonstration video was excluded from the authoring time measure. Instead, we believe it points to learning effects related to the interface and interaction workflow. This interpretation is further supported by the variability in performance. Manual authoring was not only slower on average but also less predictable across participants, with a markedly higher spread in authoring times (SD = 10 min 50 s vs. 5 min 26 s for AI-assisted authoring; approximately four times the variance), suggesting that participants benefited unevenly from familiarization with the interaction flow. In addition, experiencing AI-assisted authoring first may have shaped participants' expectations, leading to a more critical evaluation of manual authoring, reflected in lower perceived usability and higher perceived task load when experienced in that order.

6.1.2 Instruction Quality Across Conditions. While we did not experimentally validate authored instructions as part of this study, we stored them in reloadable file formats and examined them retrospectively. Based on that, it appears to us that AI assistance mainly impacted consistency rather than substantially changing the functional adequacy of authored instructions. All instructions were sufficiently descriptive for the task to be performed, independent of authoring approach. Notably, a moderately complex task was deliberately chosen, and end-users would have all authoring modalities at their disposal during guidance, whether instructions were authored manually or assisted by AI. While video instructions did not differ meaningfully between conditions, the text instructions revealed a contrast in uniformity. Manual entries were more heterogeneous, characterized by more frequent linguistic errors (29 in total) and greater fluctuations in character length. When averaging the variability found within each individual step, manual instructions proved far less predictable than the AI-assisted text (mean SD = 40.54 vs. 20.84). This trend toward standardization in the AI condition was further reflected in the higher semantic similarity of textual instructions between users for the same steps (mean cosine similarity 0.324 vs. 0.167), suggesting that the AI guides participants toward a more cohesive instructional style. The 3D instructions show a similar pattern: AI-assisted authoring naturally produced a consistent mapping of actions and 3D representations, while manual authoring led to a wider variety in 3D object compositions. Averaging over all manually authored steps across all participants, 9.36 different 3D object combinations were used per step, meaning the same task step was, on average, presented in more than nine different ways. Moreover, this does not even yet consider the spatial variance of where 3D objects were positioned. Interestingly, this can go as far as authors sometimes purposely placing 3D instructions next to their physical counterparts (e.g. circular arrow left to the rotary switch) instead of overlaying them (P_6).

From an industrial perspective, representational consistency is a highly relevant quality characteristic of work instructions [25]. While manual authoring allows for greater expressive flexibility, this flexibility can lead to divergence, especially once multiple different authors contribute to the same instruction set, resulting in inconsistencies in wording and visual representation. In contrast, AI-assisted suggestions act as a shared reference that would promote a more uniform style across contributors. At the same time, this does not mean consistency and flexibility need to be mutually exclusive. Rather, AI suggestions can provide a consistent baseline that authors adapt where necessary.

Beyond consistency, the question of whether AI-generated outputs were critically inspected needs to be asked since the risk of over-trust in AI-generated output constitutes a well-documented concern [29]. The observed correction patterns suggest that participants did engage with the generated content rather than accepting it uncritically. In total, six demonstrations were misclassified and addressed through re-demonstrations. Textual errors were similarly corrected, e.g., with the step involving closing of the lateral cell door showing the highest correction rate due to frequent confusion with opening it. Nevertheless, the risk of over-trust in generative outputs remains a considerable issue. Potential safeguards to combat this issue could involve mandatory confirmation steps and per-step checklists that prompt authors to actively verify instructions.

6.1.3 Outlier Analysis. An inspection of participant-level deviations suggests that the main pattern was robust, with exploratory outlier screening (Tukey's $1.5 \times IQR$ rule on participant-level paired differences) identifying only one formal outlier: P_{16} on SUS, with a difference score of -47.5 falling below the lower fence of -20.0. No Tukey outliers were detected for authoring time (fences: -2565.63 to 519.34 s), total trial time (fences: -3014.84 to 1011.96 s), rTLX (fences: -33.33 to 11.11), or questions asked (up to 7). This case is most plausibly explained by an erroneous response pattern, as the participant submitted a uniform rating of 5 across all SUS items, which yields a score of 50 given the alternating scoring direction of the scale. Notably, this session followed the manual-first order, lasted about 90 min, and was conducted in the evening, which may have contributed to fatigue-driven responding. The remaining data from this participant includes faster completion, fewer questions, and a stated preference for AI-assisted authoring.

One further notable case was P_{17}, the only participant who completed the manual condition faster than the AI-assisted condition. Given that this participant started with AI-assisted authoring yet still preferred it and rated it as more usable and less demanding, a carry-over or learning effect benefiting the second condition is the most plausible explanation. For perceived task load, three participants reported slightly higher rTLX scores in the AI-assisted condition, though none were flagged as statistical outliers. Taken together, these cases point to localized mismatches between subjective experience and objective efficiency for a small number of participants, rather than evidence against the overall pattern.

6.1.4 Study Limitations. In our study we deliberately evaluate against a manual baseline, closely mimicking Microsoft Dynamics 365 Guides, aiming to reflect a workflow previously used in industry. Alternative or more recent authoring tools, both manual and AI-assisted, might perform better and yield different results, making further comparative explorations worthwhile. Related to that, authorable AR constructs in the AI-assisted authoring tool VisAARD are deliberately close to what could be manually created to allow fair comparison. Yet, these might not necessarily be the ideal presentation modalities, as research has explored different representations of instructional information on AR HMDs and showed relying more on in-situ 3D modalities might be beneficial [8]. Apart from that, though we found highly significant results with large effect sizes, our sample size is relatively small and demographically homogeneous, consisting primarily of young participants from a technical university context. Prior work also suggests that AR authoring performance varies with users' digital literacy and experience [6], which may limit generalizability. In addition, just over half of our participants (52.4%) had no prior experience with the specific robotic cell, so claims about expert industrial authoring would require validation with domain experts such as technical writers or trained operators. No multiplicity correction was applied, as analyses were primarily exploratory. Under a conservative Bonferroni correction across the five main outcomes ($\alpha = 0.01$), the effects for authoring time and perceived task load remain significant, whereas the difference in perceived usability would not. A further limitation concerns our task-load measure. We assessed perceived task load with a non-validated 7-point Likert adaptation of the raw NASA-TLX instead of its validated 21-point bipolar response format, a type of modification recently criticized in HCI [2]. We chose a 7-point Likert format as a pragmatic compromise: it retained more response granularity than a five-point item scale while keeping the post-condition questionnaire in a familiar Likert-style response format rather than switching to the original NASA-TLX's distinct bipolar rating format. This adaptation is not psychometrically validated and limits comparability with standard NASA-TLX scores and with values reported in other studies. We therefore interpret the rTLX results only as within-study evidence that AI-assisted authoring was perceived as less demanding, rather than as absolute or normative workload estimates. Finally, the thematic analysis of open-ended questionnaire responses was conducted by a single researcher, which limits the reliability of the qualitative findings.

6.2 Implications for AI-Assisted AR Authoring

Having discussed the concrete study results, we now turn to broader implications for AI-assisted authoring tools in view of meaningful balance between human intervention and AI-based automation, and briefly go into promising aspects to explore in future research.

6.2.1 Toward Hybrid Authoring Workflows. The preferences of the study participants indicate that AI assistance is very much welcomed in AR instruction authoring (preferred by 20/21). Still, looking at qualitative feedback reveals that users valued different aspects of *both* manual and AI-assisted authoring approaches, each having their distinct strengths and weaknesses. More specifically, AI-assisted authoring was preferred primarily because it substantially improved efficiency (task completion time, perceived task load, and perceived usability). At an abstract level, manual authoring emphasized control and flexibility, whereas AI-assisted authoring provided structure and accelerated progress. This contrast is also reflected in concrete interaction costs; manual authoring was perceived as cumbersome due to repeated 3D manipulation, frequent UI interactions, and the need to construct every element from scratch, while AI-assisted authoring reduced this overhead and offered a useful point of reference. At the same time, participants expressed a clear desire for more flexibility to adapt system-generated instructions, for example, by adding or deleting elements. Bridging quantitative and qualitative data, this suggests that the two approaches are not competing alternatives but complementary, pointing toward a hybrid solution in which AI-assisted authoring serves as a foundation and manual customization provides targeted control. In this sense, users do not seek unrestricted freedom at all times, but agency where it matters most.

6.2.2 The Human Role in Instruction Authoring. Current and near-future AI-based assistance does not remove the human from the authoring process, particularly in the context of industrial use cases. While modern VLMs are designed for broad "reasoning", applying them in industry requires contextual understanding beyond what can be expected from pretrained models alone. At the same time, training models for every specific use case is impractical, and fine-tuning is arguably still insufficient. As a result, human intervention remains essential for producing practically usable AR instructions. Even as AI capabilities continue to improve, the need for human oversight is unlikely to disappear due to requirements such as safety and compliance. Consequently, even if future systems achieve near-perfect automatic instruction generation, human involvement will remain critical in our view. However, it appears that the role of the human in that regard could shift. Rather than acting primarily as the creator of instructions, authors increasingly take on the role of the reviewer. Based on our observations during the AI-assisted condition, participants did not simply accept the generated output; they inspected it, decided whether modifications were necessary, and performed them if so. This is also visible in average remanipulations of steps (M = 52.0%, SD = 13.8%), text edits (48.9%, SD = 23.2%), and rearrangements of panels (56.7%, SD = 37.1%) across participants during the review and refinement of AI-proposed instructions. Considering this, the primary effort shifts away from initial construction toward verification and adaptation. This aligns with participants' feedback on the particular intricateness of 3D manipulations as a major source of frustration during manual instruction authoring ($P_{10,14,16,17,19,20}$), underlining this shift as positive.

6.2.3 Technical Limitations and Future Directions. The current system follows a two-phase authoring procedure, partly motivated by the processing time VLMs require to analyze visual input, particularly at higher resolution. This design mitigates waiting times in the general workflow, though short delays remain when authors recapture and regenerate instructions during review. Exploring alternative approaches, a potential next step could be to evaluate the advantages and drawbacks of authoring based on an uninterrupted workflow demonstration against our stepwise approach. This introduces new challenges such as the temporal localization of unknown actions, an ongoing topic in research [50].

The current system positions 3D instructions relative to tracked hand poses, which means it does not support dynamic tasks requiring anchoring relative to moving objects in the scene. This connects to the broader challenge of object identification and tracking in industrial settings, where the non-generic and highly variable nature of components continues to pose a significant barrier to practical application [16]. Combining scene understanding with hand-tracking information for the placement logic could improve the contextual adequacy of 3D instruction positioning. At the same time, first research regarding the spatial arrangement of holograms based on unknown objects points in a possible direction for extending such approaches toward real-time object tracking [44]. Improving and evaluating them in the industrial context represents a necessary research avenue to ultimately achieve dynamic AR instructions.

Static 3D instructions in the current AI-assisted implementation rely on a predefined object library mapped to action types. Although unknown actions are automatically modeled using 3D hand motions, extending this library may be desirable for certain action types. Right now, this would require programming intervention, though a no-code approach is already feasible and worth investigating. One way this could be achieved would be to allow authors to define new action-instruction mappings directly in AR by selecting and positioning desired 3D objects relative to their hand poses. Newly added mappings are then automatically adopted into the system prompt, completing the no-code extension of available instruction types.

The HL2 remains the most mature optical see-through HMD available for research, e.g., due to well-established documentation and accessible development interfaces, that made it the natural choice for this work. While the study's findings do not specifically rely on used hardware and should be replicable with other optical see-through AR HMDs, we acknowledge that Microsoft has commercially discontinued the HoloLens 2, with developer support ending December 31, 2027, which carries implications for long-term deployment.

Beyond system-level improvements, we see further open questions regarding VisAARD's applicability and impact. Our focus on an industrial use case was deliberate, as it motivates an approach between human and AI that mitigates drawbacks of purely manual or automated approaches. That said, VisAARD is not inherently limited to the industrial domain; everyday procedural workflows such as healthcare and nursing procedures, laboratory protocols, or construction tasks present equally viable applications which would need to be explored. In either case, explicitly evaluating the quality of resulting instructions is important since the usefulness of an AR authoring tool is ultimately determined by both the usability of the tool itself and the usefulness of the AR instructions it produces [4].

While this study examined only the former empirically, we plan to evaluate the latter through expert reviews in a next step.

In the broader context of AI-assisted authoring, we furthermore see particular potential for multimodal AI to support the creation of not only useful but also cognitively accessible instructions. Incorporating AI into the authoring process offers the unique opportunity to consider accessibility from the outset rather than as an afterthought [30]. The automatic creation of instructions at different levels of detail, e.g., adapted to different levels of expertise (a feature also explicitly requested by participants $P_{3,7,10}$), could be another direction to investigate, allowing the same authored content to meet varying user needs without additional authoring effort.

7 Conclusion

Authoring AR-based procedural instructions remains a major bottleneck for the practical adoption of AR guidance. In this paper, we presented VisAARD, an open-source, human-in-the-loop AR authoring tool that uses a vision-language model to interpret stepwise demonstrations and generate draft AR instructions. In a within-subject study with 21 participants, we compared VisAARD against a manual authoring baseline modeled after the industry-standard Microsoft Dynamics 365 Guides. VisAARD substantially reduced raw authoring time, lowered perceived task load, yielded higher perceived usability ratings, and was preferred by 20 of 21 participants. While these quantitative results clearly favored AI-assisted authoring across all measured outcomes, thematically analyzed qualitative feedback indicated that participants still valued the agency and expressiveness of the manual authoring condition. Together, these findings motivate hybrid authoring workflows in which AI-generated structure serves as a foundation for targeted manual refinement. In industrial contexts, where safety and compliance require human oversight regardless of AI capability anyway, we believe that designing authoring tools around this reviewer role appears more productive than pursuing full automation.

Acknowledgments

As this paper examines AI-assisted authoring, naturally, we also used AI in preparing this manuscript, primarily to restructure text from author notes and improve readability and style. In line with the human-in-the-loop perspective discussed in this paper, all data collection, analysis, discussion, and interpretation were carried out by the authors. This research was funded by the Ministry of Science, Research and the Arts of Baden-Württemberg (MWK) (grant BW6_03). Jonas Blattgerste is funded by zukunft.niedersachsen, a joint science funding program of the Lower Saxony Ministry of Science and Culture and the Volkswagen Foundation.

References

[1] Ilker Baki Alkan and Hasan Basri Basaga. 2023. Augmented reality technologies in construction project assembly phases. *Automation in Construction* 156 (Dec. 2023), 105107. doi:10.1016/j.autcon.2023.105107

[2] Ebrahim Babaei, Tilman Dingler, Benjamin Tag, and Eduardo Velloso. 2025. Should we use the NASA-TLX in HCI? A review of theoretical and methodological issues around mental workload measurement. *International Journal of Human-Computer Studies* 201 (2025), 103515. doi:10.1016/j.ijhcs.2025.103515

[3] Bhaskar Bhattacharya and Eliot H. Winer. 2019. Augmented reality via expert demonstration authoring (AREDA). *Computers in Industry* 105 (Feb. 2019), 61–79. doi:10.1016/j.compind.2018.04.021

[4] Jonas Blattgerste. 2024. *The design space of augmented reality authoring tools and its exploration for the procedural training context.* Ph. D. Dissertation. Universität Bielefeld, Bielefeld, Germany. doi:10.4119/unibi/2987473

[5] Jonas Blattgerste, Jan Behrends, and Thies Pfeiffer. 2022. A web-based analysis toolkit for the system usability scale. In *Proceedings of the 15th International Conference on PErvasive Technologies Related to Assistive Environments* (Corfu, Greece) *(PETRA '22)*. Association for Computing Machinery, New York, NY, USA, 237–246. doi:10.1145/3529190.3529216

[6] Jonas Blattgerste, Jan Behrends, and Thies Pfeiffer. 2023. TrainAR: An open-source visual scripting-based authoring tool for procedural mobile augmented reality trainings. *Information* 14, 4 (April 2023), 219. doi:10.3390/info14040219

[7] Jonas Blattgerste, Patrick Renner, and Thies Pfeiffer. 2019. Authorable augmented reality instructions for assistance and training in work environments. In *Proceedings of the 18th International Conference on Mobile and Ubiquitous Multimedia (MUM '19)*. Association for Computing Machinery, New York, NY, USA, 1–11. doi:10.1145/3365610.3365646

[8] Jonas Blattgerste, Patrick Renner, Benjamin Strenge, and Thies Pfeiffer. 2018. In-situ instructions exceed side-by-side instructions in augmented reality assisted assembly. In *Proceedings of the 11th PErvasive Technologies Related to Assistive Environments Conference (PETRA '18)*. Association for Computing Machinery, New York, NY, USA, 133–140. doi:10.1145/3197768.3197778

[9] Virginia Braun and Victoria Clarke. 2019. Reflecting on reflexive thematic analysis. *Qualitative research in sport, exercise and health* 11, 4 (2019), 589–597. doi:10.1080/2159676X.2019.1628806

[10] John Brooke. 1996. SUS: A 'quick and dirty' usability scale. In *Usability Evaluation in Industry*. Taylor & Francis, London, 189–194.

[11] Pierre Bégout, Thierry Duval, Sébastien Kubicki, Bertrand Charbonnier, and Emmanuel Bricard. 2020. WAAT: A workstation AR authoring tool for Industry 4.0. In *Augmented Reality, Virtual Reality, and Computer Graphics*. Springer International Publishing, Cham, 304–320. doi:10.1007/978-3-030-58468-9_22

[12] Subramanian Chidambaram, Sai Swarup Reddy, Matthew Rumple, Ananya Ipsita, Ana Villanueva, Thomas Redick, Wolfgang Stuerzlinger, and Karthik Ramani. 2022. EditAR: A digital twin authoring environment for creation of AR/VR and video instructions from a single demonstration. In *2022 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*. IEEE, 326–335. doi:10.1109/ISMAR55827.2022.00048

[13] Lea M. Daling and Sabine J. Schlittmeier. 2024. Effects of augmented reality-, virtual reality-, and mixed reality–based training on objective performance measures and subjective evaluations in manual assembly tasks: A scoping review. *Human Factors* 66, 2 (Feb. 2024), 589–626. doi:10.1177/00187208221105135

[14] Luís Fernando de Souza Cardoso, Flávia Cristina Martins Queiroz Mariano, and Ezequiel Roberto Zorzal. 2020. A survey of industrial augmented reality. *Computers & Industrial Engineering* 139 (Jan. 2020), 106159. doi:10.1016/j.cie.2019.106159

[15] Wenjing Deng, Sisi He, Xintong Wu, Yiran Wan, Chenzhuo Xiang, Zhihao Yao, and Haipeng Mi. 2025. AI extraction of disassembly videos for generating AR repair instructions. In *Adjunct Proceedings of the 38th Annual ACM Symposium on User Interface Software and Technology (UIST Adjunct '25)*. Association for Computing Machinery, New York, NY, USA, 1–3. doi:10.1145/3746058.3758436

[16] Jeevan S. Devagiri, Sidike Paheding, Quamar Niyaz, Xiaoli Yang, and Samantha Smith. 2022. Augmented reality and artificial intelligence in industry: trends, tools, and future challenges. *Expert Systems with Applications* 207 (Nov. 2022), 118002. doi:10.1016/j.eswa.2022.118002

[17] Juan C Dibene and Enrique Dunn. 2022. HoloLens 2 sensor streaming. doi:10.48550/arXiv.2211.02648

[18] M Eswaran and M V A Raju Bahubalendruni. 2022. Challenges and opportunities on AR/VR technologies for manufacturing systems in the context of industry 4.0: A state of the art review. *Journal of Manufacturing Systems* 65 (Oct. 2022), 260–278. doi:10.1016/j.jmsy.2022.09.016

[19] Wei Fang, Lixi Chen, Tienong Zhang, Chengjun Chen, Zhan Teng, and Lihui Wang. 2023. Head-mounted display augmented reality in manufacturing: A systematic review. *Robotics and Computer-Integrated Manufacturing* 83 (Oct. 2023), 102567. doi:10.1016/j.rcim.2023.102567

[20] Eléonore Ferrier-Barbut, Ignacio Avellino, Marie-Aude Vitrani, and Geoffroy Canlorbe. 2023. Head mounted displays in surgical training and planning: a literature review. In *Proceedings of the 34th Conference on l'Interaction Humain-Machine (IHM '23)*. Association for Computing Machinery, New York, NY, USA, 1–10. doi:10.1145/3583961.3583973

[21] Markus Funk, Lars Lischke, Sven Mayer, Alireza Sahami Shirazi, and Albrecht Schmidt. 2018. Teach me how! Interactive assembly instructions using demonstration and in-situ projection. In *Assistive Augmentation*. Springer Singapore, Singapore, 49–73. doi:10.1007/978-981-10-6404-3_4

[22] Alastair Hampshire, Hartmut Seichter, Raphaël Grasset, and Mark Billinghurst. 2006. Augmented reality authoring: generic context from programmer to designer. In *Proceedings of the 18th Australia Conference on Computer-Human Interaction: Design: Activities, Artefacts and Environments*. Association for Computing Machinery, Sydney, Australia, 409–412. doi:10.1145/1228175.1228259

[23] Sandra G. Hart. 2006. NASA-Task Load Index (NASA-TLX); 20 years later. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* 50, 9 (2006),

904–908. doi:10.1177/154193120605000909

[24] Sandra G. Hart and Lowell E. Staveland. 1988. Development of NASA-TLX (Task Load Index): results of empirical and theoretical research. In *Human Mental Workload*. Advances in Psychology, Vol. 52. North-Holland, 139–183. doi:10.1016/S0166-4115(08)62386-9

[25] Anders Haug. 2015. Work instruction quality in industrial management. *International Journal of Industrial Ergonomics* 50 (Nov. 2015), 170–177. doi:10.1016/j.ergon.2015.09.015

[26] High-Level Expert Group on Artificial Intelligence. 2019. *Ethics guidelines for trustworthy AI*. Technical Report ES-AA-19-001-EN-N. European Commission, Publications Office of the European Union, Luxembourg. doi:10.2759/346720

[27] Panayu Keelawat and Ryo Suzuki. 2024. Transforming procedural instructions into in-situ augmented reality guides with InstructAR. In *The 37th Annual ACM Symposium on User Interface Software and Technology (UIST '24)*. Association for Computing Machinery, New York, NY, USA, 1–3. doi:10.1145/3672539.3686321

[28] Philipp Klimant and Christian Kollatsch. 2022. Concepts for creating augmented reality based technical documentations for the maintenance of machine tools. *International Journal on Interactive Design and Manufacturing (IJIDeM)* 16, 2 (June 2022), 765–773. doi:10.1007/s12008-022-00880-7

[29] Artur Klingbeil, Cassandra Grützner, and Philipp Schreck. 2024. Trust and reliance on AI — An experimental study on the extent and costs of overreliance on AI. *Computers in Human Behavior* 160 (Nov. 2024), 108352. doi:10.1016/j.chb.2024.108352

[30] Valentin Knoben, Jonas Blattgerste, Thies Pfeiffer, Björn Hein, and Christian Wurl. 2026. Towards augmented reality authoring tools that ensure cognitive accessibility of instructions. In *2026 IEEE International Conference on Artificial Intelligence and eXtended and Virtual Reality (AIxVR)*. 345–352. doi:10.1109/AIxVR67263.2026.00060

[31] Junhan Kong, Dena Sabha, Jeffrey P. Bigham, Amy Pavel, and Anhong Guo. 2021. TutorialLens: authoring interactive augmented reality tutorials through narration and demonstration. In *Symposium on Spatial User Interaction (SUI '21)*. Association for Computing Machinery, New York, NY, USA, 1–11. doi:10.1145/3485279.3485289

[32] Thomas Kosch, Pascal Knierim, Mareike Kritzler, Daniel Beicht, and Florian Michahelles. 2023. Lenssembly: authoring assembly instructions in augmented reality using programming-by-demonstration. In *Human-Technology Interaction: Shaping the Future of Industrial User Interfaces*. Springer International Publishing, Cham, 199–222. doi:10.1007/978-3-030-99235-4_8

[33] Traian Lavric, Emmanuel Bricard, Marius Preda, and Titus Zaharia. 2020. An AR work instructions authoring tool for human-operated industrial assembly lines. In *2020 IEEE International Conference on Artificial Intelligence and Virtual Reality (AIVR)*. IEEE, Utrecht, Netherlands, 174–183. doi:10.1109/AIVR50618.2020.00037

[34] Jaewook Lee, Filippo Aleotti, Diego Mazala, Guillermo Garcia-Hernando, Sara Vicente, Oliver James Johnston, Isabel Kraus-Liang, Jakub Powierza, Donghoon Shin, Jon E. Froehlich, Gabriel Brostow, and Jessica Van Brummelen. 2025. ImaginateAR: AI-assisted in-situ authoring in augmented reality. In *Proceedings of the 38th Annual ACM Symposium on User Interface Software and Technology (UIST '25)*. Association for Computing Machinery, New York, NY, USA, 1–21. doi:10.1145/3746059.3747635

[35] Junjian Lin, Jianjian Wang, Pingfa Feng, Xiangyu Zhang, Dingwen Yu, and Jianfu Zhang. 2025. AI-aided automated AR-assisted assembly instruction authoring and generation method. *Journal of Manufacturing Systems* 83 (Dec. 2025), 405–423. doi:10.1016/j.jmsy.2025.08.019

[36] Ziyi Liu, Zhengzhe Zhu, Enze Jiang, Feichi Huang, Ana M. Villanueva, Xun Qian, Tianyi Wang, and Karthik Ramani. 2023. InstruMentAR: auto-generation of augmented reality tutorials for operating digital instruments through recording embodied demonstration. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems (CHI '23)*. Association for Computing Machinery, New York, NY, USA, 1–17. doi:10.1145/3544548.3581442

[37] Yuqian Lu, Hao Zheng, Saahil Chand, Wanqing Xia, Zengkun Liu, Xun Xu, Lihui Wang, Zhaojun Qin, and Jinsong Bao. 2022. Outlook on human-centric manufacturing towards Industry 5.0. *Journal of Manufacturing Systems* 62 (Jan. 2022), 612–627. doi:10.1016/j.jmsy.2022.02.001

[38] Microsoft. 2026. Dynamics 365 Mixed Reality. https://learn.microsoft.com/en-us/dynamics365/mixed-reality/guides/overview (accessed 2026-01-15).

[39] Mixed Reality Toolkit Organization. 2025. MixedRealityToolkit-Unity. https://github.com/MixedRealityToolkit/MixedRealityToolkit-Unity release core-v3.3.0.

[40] Michael Nebeling and Maximilian Speicher. 2018. The trouble with augmented reality/virtual reality authoring tools. In *2018 IEEE International Symposium on Mixed and Augmented Reality Adjunct (ISMAR-Adjunct)*. IEEE, Munich, Germany, 333–337. doi:10.1109/ISMAR-Adjunct.2018.00098

[41] OpenAI. 2025. Introducing GPT-5.2. https://openai.com/index/introducing-gpt-5-2/

[42] Jeffrey N. Rouder, Paul L. Speckman, Dongchu Sun, Richard D. Morey, and Geoffrey Iverson. 2009. Bayesian t tests for accepting and rejecting the null hypothesis. *Psychonomic Bulletin & Review* 16, 2 (2009), 225–237. doi:10.3758/PBR.16.2.225

[43] Joel Murithi Runji, Yun-Ju Lee, and Chih-Hsing Chu. 2023. Systematic literature review on augmented reality-based maintenance applications in manufacturing centered on operator needs. *International Journal of Precision Engineering and Manufacturing-Green Technology* 10, 2 (March 2023), 567–585. doi:10.1007/s40684-022-00444-w

[44] Aditya Sharma, Luke Yoffe, and Tobias Höllerer. 2024. OCTO+: a suite for automatic open-vocabulary object placement in mixed reality. In *2024 IEEE International Conference on Artificial Intelligence and eXtended and Virtual Reality (AIxVR)*. 157–165. doi:10.1109/AIxVR59861.2024.00028

[45] Jingyu Shi, Rahul Jain, Seunggeun Chi, Hyungjun Doh, Hyung-gun Chi, Alexander J. Quinn, and Karthik Ramani. 2025. CARING-AI: towards authoring context-aware augmented reality instruction through generative artificial intelligence. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems (CHI '25)*. Association for Computing Machinery, New York, NY, USA, 1–23. doi:10.1145/3706598.3713348

[46] Lucchas Ribeiro Skreinig, Peter Mohr, Blanca Berger, Markus Tatzgern, Dieter Schmalstieg, and Denis Kalkofen. 2024. Immersive authoring by demonstration of industrial procedures. In *2024 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*. IEEE. doi:10.1109/ISMAR62088.2024.00146

[47] Ana Stanescu, Peter Mohr, Dieter Schmalstieg, and Denis Kalkofen. 2022. Model-free authoring by demonstration of assembly instructions in augmented reality. *IEEE Transactions on Visualization and Computer Graphics* 28, 11 (Nov. 2022), 3821–3831. doi:10.1109/TVCG.2022.3203104

[48] Daniel Stover and Doug Bowman. 2024. TAGGAR: general-purpose task guidance from natural language in augmented reality using vision-language models. In *ACM Symposium on Spatial User Interaction (SUI '24)*. Association for Computing Machinery, New York, NY, USA, 1–12. doi:10.1145/3677386.3682095

[49] Eric-Jan Wagenmakers, Jonathon Love, Maarten Marsman, Tahira Jamil, Alexander Ly, Josine Verhagen, Ravi Selker, Quentin F. Gronau, Damian Dropmann, Bruno Boutin, Frans Meerhoff, Patrick Knight, Akash Raj, Erik-Jan van Kesteren, Johnny van Doorn, Martin Šmíra, Sacha Epskamp, Alexander Etz, Dora Matzke, Tim de Jong, Don van den Bergh, Alexandra Sarafoglou, Helen Steingroever, Koen Derks, Jeffrey N. Rouder, and Richard D. Morey. 2018. Bayesian inference for psychology. Part II: Example applications with JASP. *Psychonomic Bulletin & Review* 25, 1 (2018), 58–76. doi:10.3758/s13423-017-1323-7

[50] Naoki Wake, Atsushi Kanehira, Kazuhiro Sasabuchi, Jun Takamatsu, and Katsushi Ikeuchi. 2024. Open-vocabulary action localization with iterative visual prompting. doi:10.48550/arXiv.2408.17422

[51] Tracey L Weissgerber, Natasa M Milic, Stacey J Winham, and Vesna D Garovic. 2015. Beyond bar and line graphs: time for a new data presentation paradigm. *PLOS Biology* 13, 4 (2015), e1002128. doi:10.1371/journal.pbio.1002128

[52] Matt Whitlock, George Fitzmaurice, Tovi Grossman, and Justin Matejka. 2020. AuthAR: concurrent authoring of tutorials for AR assembly guidance. In *Proceedings of Graphics Interface 2020*. Canadian Human-Computer Communications Society, 431–439. doi:10.20380/GI2020.43

[53] Masahiro Yamaguchi, Shohei Mori, Peter Mohr, Markus Tatzgern, Ana Stanescu, Hideo Saito, and Denis Kalkofen. 2020. Video-annotated augmented reality assembly tutorials. In *Proceedings of the 33rd Annual ACM Symposium on User Interface Software and Technology (UIST '20)*. Association for Computing Machinery, New York, NY, USA, 1010–1022. doi:10.1145/3379337.3415819

[54] Marcus Yung, Ahmet Kolus, Richard Wells, and W. Patrick Neumann. 2020. Examining the fatigue-quality relationship in manufacturing. *Applied Ergonomics* 82 (Jan. 2020), 102919. doi:10.1016/j.apergo.2019.102919

[55] Ada Yi Zhao, Aditya Gunturu, Ellen Yi-Luen Do, and Ryo Suzuki. 2025. Guided reality: generating visually-enriched AR task guidance with LLMs and vision models. In *Proceedings of the 38th Annual ACM Symposium on User Interface Software and Technology (UIST '25)*. Association for Computing Machinery, New York, NY, USA, 1–15. doi:10.1145/3746059.3747784

[56] Chenfei Zhu, Shao-Kang Hsia, Xiyun Hu, Ziyi Liu, Jingyu Shi, and Karthik Ramani. 2025. AgentAR: creating augmented reality applications with tool-augmented LLM-based autonomous agents. In *Proceedings of the 38th Annual ACM Symposium on User Interface Software and Technology (UIST '25)*. Association for Computing Machinery, New York, NY, USA, 1–23. doi:10.1145/3746059.3747676

[57] J. Zhu, S. K. Ong, and A. Y. C. Nee. 2013. An authorable context-aware augmented reality system to assist the maintenance technicians. *The International Journal of Advanced Manufacturing Technology* 66 (Aug. 2013), 1699–1714. doi:10.1007/s00170-012-4451-2