From Structure to Semantics: Hypergraph-Based AR Assembly Guidance with LLM-Mediated Narration

Xinda Liu, Bowei Zhang, Jiaju Xu, Jian Wu*, Guohua Geng, and Lili Wang

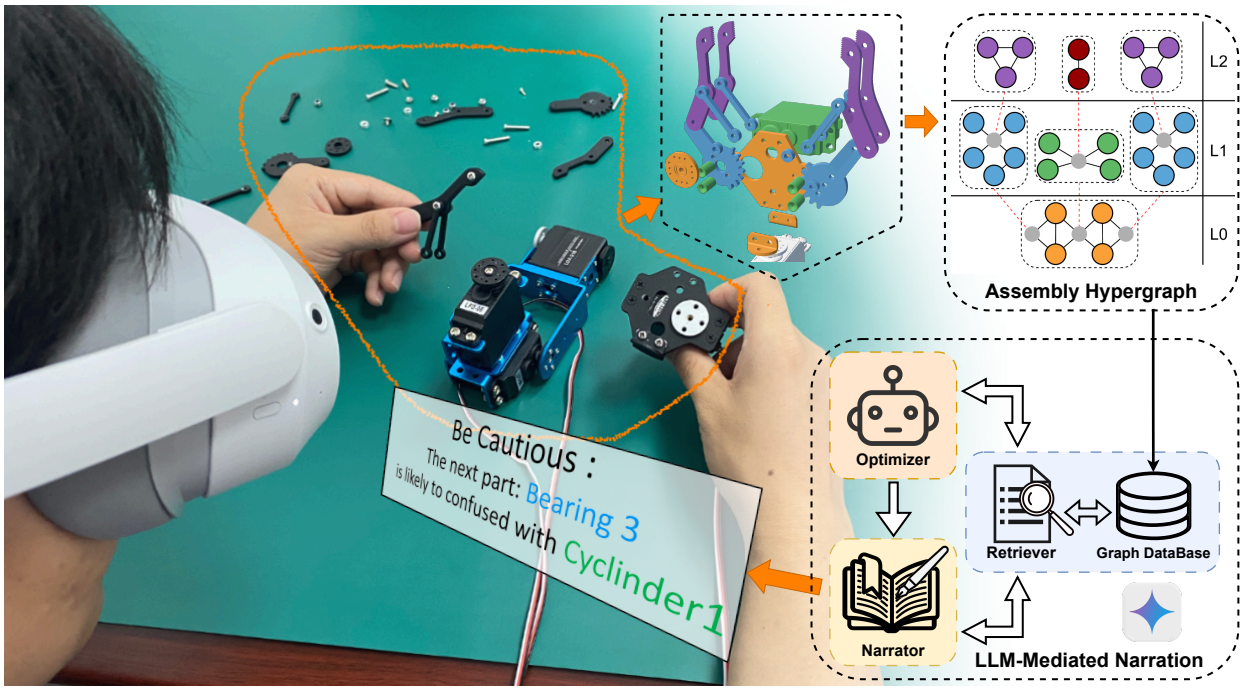

Fig. 1: **Assembly Hypergraph with LLM-Mediated Narration and Augmented Reality (AR) guidance.** We design an LLM-mediated narration method to fuse semantic information and human-centric optimization in the assembly hypergraph. Our method can translate the tedious assembly sequence into a narration polished by LLM narrators, improving both task performance and user experience.

Abstract— Effective Augmented Reality (AR) guidance for complex assembly faces a dual challenge: the inability of conventional liaison graphs to represent procedural logic, and the cognitive burden imposed by visual instructions. We argue that the solution requires a more expressive structure to overcome these representational deficits and a narration approach to mediate instruction complexity. Our method first employs an assembly hypergraph to capture the task's hierarchical information, from which an A* search algorithm generates an optimal assembly path. Then a Large Language Model (LLM)-mediated narration workflow is designed to address the ergonomic deficiencies of the machine-centric path. It employs an optimizer to improve fluency, followed by a narrator that crafts the steps into an intuitive instruction narration. A within-subjects user study ($N = 24$) revealed a progressive enhancement from our method's components. The transition from a liaison-graph baseline to the hypergraph alone improved objective outcomes by reducing task time and errors and improving subjective ratings (SUS, NASA-TLX, TAM, and ARI). Subsequently, augmenting the LLM-mediated narration maintained these gains while lowering cognitive load and elevating user experience and usability. Our findings indicate the value of our AR assembly design and discuss the opportunities of using LLM as a mediation layer for better user interaction.

Index Terms—Augmented Reality, Assembly Guidance, Hypergraphs, Large Language Model (LLM).

1 Introduction

Augmented reality (AR) has emerged as a powerful tool in enhancing industrial assembly processes, offering substantial potential to improve both production efficiency and human performance [41, 55, 59]. However, the effectiveness of AR guidance is fundamentally constrained by its ability to accurately represent the structure and state of the assembly task [26, 58]. Such systems often struggle to model the complex physical and logical relationships among numerous components, which in turn compromises the quality and reliability of the guidance provided [10, 32].

The prevailing approach to modeling assembly tasks relies on liaison graphs [19, 39, 53], in which nodes represent individual parts and edges correspond to direct physical connections between them. This model offers a clear advantage due to its straightforward one-to-one mapping with the physical assembly, facilitating intuitive authoring and interpretation for many conventional procedures. However, a significant limitation arises when attempting to encode procedural logic rather

* *Corresponding author*
- *Xinda Liu, Bowei Zhang, Jiaju Xu, Guohua Geng are with Northwest University, 710127, Xi'an, China. E-mail: {liuxinda, ghgeng}@nwu.edu.cn, {202322922, xujiaju}@stumail.nwu.edu.cn*
- *Jian Wu, Lili Wang are with Beihang University, 100191, Beijing, China. E-mail: {lanayawj, wanglily}@buaa.edu.cn*

Manuscript received xx xxx. 201x; accepted xx xxx. 201x. Date of Publication xx xxx. 201x; date of current version xx xxx. 201x. For information on obtaining reprints of this article, please send e-mail to: reprints@ieee.org. Digital Object Identifier: xx.xxxx/TVCG.201x.xxxxxxx

than merely the final spatial configuration of the assembly process [67]. Because liaison graphs inherently capture spatial relationships among components, they are poorly suited to representing higher-level assembly constructs such as sub-assemblies, multi-stage tasks, procedural dependencies, or alternative error-recovery steps [64]. This representational gap critically impedes the capacity of the system to provide consistent guidance and ensure stable transitions between assembly phases, highlighting the need for more expressive models that can capture both structural and procedural aspects of assembly tasks.

To address the aforementioned representational limitations, we propose a hypergraph-based model for AR-assisted assembly that employs hyper-edges to capture higher-order relationships such as sub-assemblies, multi-stage processes, and their associated constraints, which go beyond binary physical connections. This model not only encodes connectivity but also incorporates ordering constraints, mutual exclusions, prerequisites, and error-recovery options as fundamental elements. While this expressiveness significantly enhances system-level consistency checking and validation, it introduces substantial challenges at the user interface level, particularly the risk of high cognitive load, increased context switching, and reduced immersion when presenting such a complex representation directly in head-mounted displays [48, 57]. To bridge this gap, we present an approach that transforms the formal assembly plan derived from the hypergraph into instructions aligned with human cognitive and operational patterns. The proposed method first applies an A* search algorithm to the hypergraph to generate a structurally optimal and procedurally sound assembly path that satisfies all constraints, yet often remains unsuitable for human execution due to its rigidity. To overcome this limitation, we further propose a workflow coupling Large Language Model (LLM) with an assembly hypergraph, which performs semantic analysis and ergonomic optimization on the task sequence. This workflow reinterprets the machine-oriented plan, regroups and re-orders instructions to enhance procedural coherence, and ultimately delivers a refined guidance sequence that minimizes cognitive overhead by reducing context switches and aligning step granularity with the natural hierarchy of the assembly task.

Our main contributions are summarized as follows:

(1) We introduce a novel hypergraph-based representation that explicitly models complex assembly structures and constraints, enabling the encapsulation of higher-order relationships such as sub-assemblies, multi-stage sequences, and procedural dependencies;

(2) We propose LLM-mediated narration guidance, which structurally and semantically reorganizes task sequences to translate assembly paths into ergonomically organized task plans, thereby reducing cognitive load and enhancing operational fluency;

(3) We conduct a within-subjects user study to empirically evaluate the proposed approach against a conventional liaison graph baseline and a hypergraph-only condition. The study assesses both objective performance metrics and subjective user experience, providing comprehensive evidence for the effectiveness of integrating structured knowledge representation with LLM-mediated narration guidance.

2 Related Work

2.1 Augmented Reality Assembly

The use of AR guidance to support industrial assembly is a well-established research area [5, 19, 27, 47]. The domain is characterized by tasks that involve numerous components, intricate dependencies, and staged progressions [43]. The foundational works in AR authoring have long emphasized the importance of hierarchical structures to organize and convey complex assembly instructions. More recent approaches have focused on the dynamic aspects of the process, with a significant emphasis on state-aware systems that can detect the current configuration and validate transitions to subsequent legal states [50]. Regarding standardized visual guidance in AR, existing empirical studies remain limited and inconclusive [36, 61]. Although some research suggests that it can effectively reduce error rates or shorten completion times, systematic validation is still lacking [35, 45]. Moreover, objective performance improvements do not necessarily align with reductions in users' subjective cognitive load, suggesting a persistent gap between performing a task correctly and performing it with ease [3, 35]. In light of these gaps, this work considers how assembly guidance can be designed not only to ensure procedural correctness, but also to reduce cognitive strain during real-world operations.

2.2 Computational Task Planning and Representation

To manage the complexity of assembly tasks, the field has evolved from classical planning formalisms toward state-aware graph representations. Foundational work in AI planning established rigorous standards for logical correctness. The Planning Domain Definition Language (PDDL) provides a syntax for defining action sequences and verifying plan feasibility [2], while Hierarchical Task Networks (HTN) effectively model the decomposition of high-level goals into executable sub-tasks [16]. For systems requiring concurrency management, Petri Nets offer robust tools for modeling asynchronous flows in manufacturing environments [22]. Despite these strengths in logical verification, symbolic planners often encounter difficulties when considering the ambiguity of human interaction [4, 11]. Consequently, AR researchers have increasingly adopted graph-based representations to explicitly model state changes and enforce correctness in dynamic physical settings [33, 39, 50, 53, 60, 65, 66]. Graph structures enhance the robustness of object tracking [23, 39, 50] and define the topology of valid state transitions [53, 60]. However, the conventional liaison graph, which maps parts to nodes and physical connections to edges, lacks the semantic capacity to represent higher-level constructs such as multi-part constraints or functional sub-assemblies. Motivated by these limitations, our work adopts a Hypergraph representation that integrates the rigor of planning logic with the semantic expressiveness required for AR guidance.

2.3 Cognitive-Aware Task Narrativization in XR

No matter how robust the internal task model is, it remains inert until effectively conveyed to the user. A central difficulty in this communication process lies in managing user cognitive load [3, 49]. From the perspective of Cognitive Load Theory, AR systems must navigate the Split-Attention Effect, which posits that performance degrades when users' integrate spatially separated information sources [46]. While AR employs the Spatial Contiguity Principle to mitigate this by overlaying instructions on physical objects [9], prolonged exposure to immersive environments can still induce fatigue and discomfort [6]. Research indicates that although exposing users to a complete causal hierarchy improves understanding, it often imposes a steep learning curve [30]. When this balance between completeness and digestibility is missed, it diminishes the user's sense of presence [37, 56] and induces frustration [62]. To objectively evaluate these impacts beyond subjective reporting, recent studies have utilized eye-tracking metrics and gaze transition for real-time cognitive load and cybersickness assessment [1, 18, 34, 52]. Moreover, prevailing authoring paradigms often lack the dynamism needed to address these cognitive constraints. Structured approaches such as the trigger–action model rely on the author's ability to anticipate all interaction paths [12]. Such static design scales poorly when faced with complex real-world tasks and lacks the flexibility to adapt to emergent user behaviors [40, 51]. Large Language Models (LLMs) offer a promising avenue for dynamic narrativization but introduce the risk of hallucination, where generated instructions may be linguistically plausible but factually incorrect with known constraints [31, 38]. To mitigate this in industrial contexts, generative outputs must be grounded in verifiable logic or physical affordances [7, 29]. Recent neurosymbolic frameworks demonstrate that coupling LLMs with symbolic planners significantly improves the reliability of task reasoning [11, 24, 44]. In response, our method considers how structured task logic can be dynamically translated via LLMs, using the hypergraph as the grounding context. This approach ensures that narration remain procedurally rigorous while being sensitive to cognitive constraints, thereby bridging the gap between system optimality and user experience.

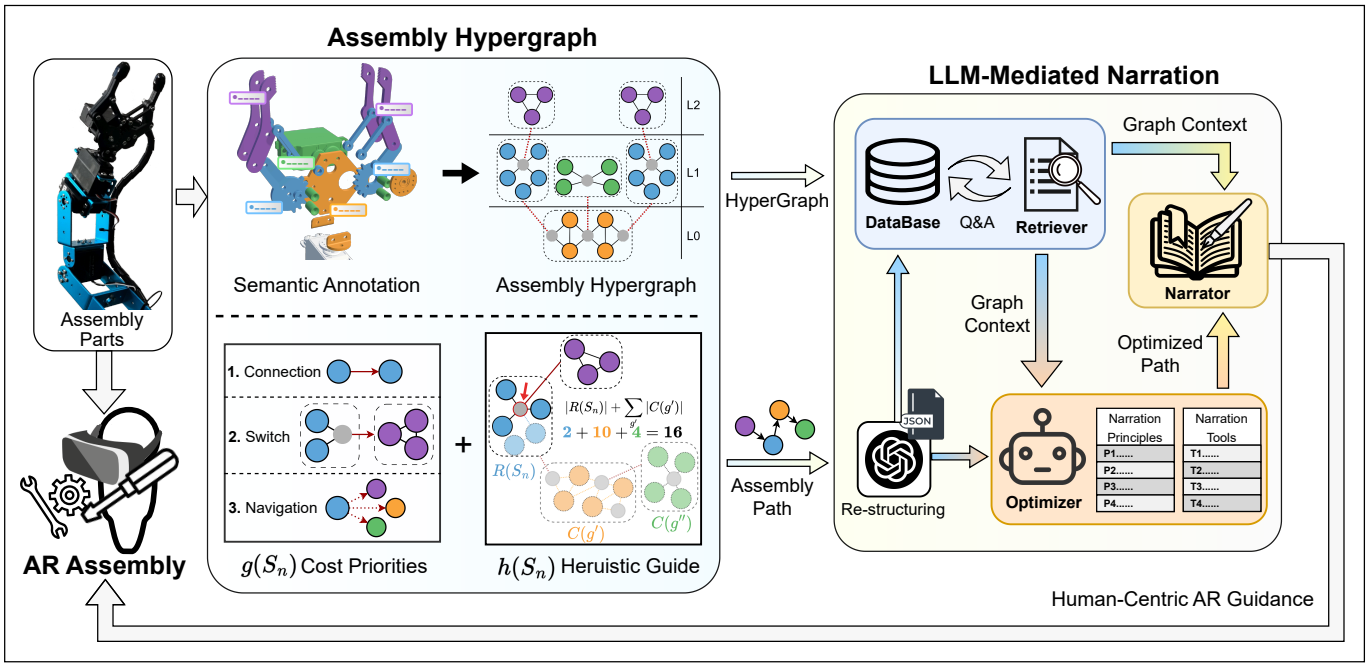

Fig. 2: **The framework of LLM-Mediated Narration based on Hypergraph.** First, assembly hypergraph is constructed with semantic information to model the hierarchical structure from raw assembly, and an A* algorithm finds an efficient path. This path then undergoes LLM-mediated narration where an **Optimizer** and **Retriever** collaborate to refine the sequence for human ergonomics using contextual data from the hypergraph. The process concludes with a **Narrator** translating the optimized path into human-centric AR guidance.

3 Method

3.1 Overview of LLM-Mediated Narration based on Assembly Hypergraph

Our method transforms a physical assembly task into human-centric AR instructions. As shown in Figure 2, the process is divided into two primary phases: first, generating a machine-optimal assembly plan from a structured representation, and second, refining this plan into an ergonomic narration.

The first phase begins by formalizing the assembly task into a semantically rich hypergraph. This structure captures not only the physical components and their connections but also the hierarchical and procedural constraints. Operating on this hypergraph, a specialized A* search algorithm generates a logically coherent assembly path which serves as the raw input for the next phase.

The second phase takes the machine-centric path and transforms it through a three-component workflow. The path is first converted into a stable data format. It then undergoes ergonomic refinement, where an **Optimizer** module re-orders steps based on human-factors principles, using a **Retriever** to query the hypergraph for context. Finally, a **Narrator** module converts the optimized path into a script of AR tool calls, producing the final guidance.

3.2 Hypergraph for Structured Assembly Representation

We adopt a hypergraph as the formal representational foundation to capture both the physical state of an assembly and the intricate logic governing its construction process. This model extends beyond conventional liaison graphs by explicitly encoding hierarchy, staging, and complex procedural constraints essential for valid guidance generation.

3.2.1 Formalization of the Assembly Hypergraph

To formally represent the assembly, we define a semantically enriched hypergraph G. In this definition, the hypergraph topology provides the abstract skeleton of connectivity, while specific mapping functions assign the physical and procedural semantics required for execution.

The foundational structure of the assembly is a triplet $G = (V, E, H)$ where V is a finite set of vertices representing fundamental entities, E is a set of standard edges representing direct physical connections, and H is a set of hyper-edges used for grouping vertices.

We then introduce the assembly semantics. We imbue the hypergraph G with assembly semantics through a family of formal maps that specify the meaning of its components. A typing function $\tau : V \to \{part, state\}$ assigns a role to each vertex. This partitions V into a set of physical parts V_{part} and abstract state nodes V_{state}. Standard edges are constrained to model connections only between physical parts, i.e., $E \subseteq \{\{u,v\} \mid u,v \in V_{part}, u \neq v\}$. An attribute map $\alpha_E : E \to \Sigma_E$ assigns physical properties to each connection. Crucially, α_E serves as an integrity constraint check. A connection $e = \{u,v\}$ is valid only if the physical properties are compatible between u and v.

Hyper-edges provide hierarchical and procedural depth compared to normal edges. The second typing function, $\sigma : H \to \{sub\text{-}assembly, stage\}$, is used to distinguish their roles. A sub-assembly hyper-edge ($h \in H_{sub}$) groups a set of part vertices, defined by an incidence map $\iota : H_{sub} \to 2^{V_{part}}$. A stage hyper-edge ($h \in H_{stage}$) delineates a distinct step in the assembly process. Each stage is associated with procedural attributes by an attribute map $\alpha_H : H_{stage} \to \Sigma_H$.

Building on the concept of stages, the procedural logic is defined by two binary relations on the set of stages H_{stage}. The prerequisite relation, $\mathrm{R_{pre}} \subseteq H_{stage} \times H_{stage}$, where $(h_i, h_j) \in \mathrm{R_{pre}}$ signifies that stage h_i must be completed before stage h_j. The mutual exclusion relation, $\mathrm{R_{ex}} \subseteq H_{stage} \times H_{stage}$, where $(h_i, h_j) \in \mathrm{R_{ex}}$ indicates that stages h_i and h_j cannot be performed concurrently. For a valid workflow, the reflexive transitive closure $\mathrm{R_{pre}}^*$ of the prerequisite relation induces a partial order $(H_{stage}, \preceq)$ on the assembly stages. In essence, the H_{stage} define the task that links the associated physical parts together.

Taken together, the structural triplet (V, E, H) and the semantic components, which include the maps $(\tau, \sigma, \alpha_E, \alpha_H, \iota)$ and relations $(\mathrm{R_{pre}}, \mathrm{R_{ex}})$, form a comprehensive hypergraph of the assembly task. This model explicitly captures the component hierarchy, procedural stages, and operational constraints, thereby establishing a robust foundation for the algorithmic planning of an assembly sequence.

3.2.2 Assembly Path Generation Via A* Search

To traverse the assembly hypergraph, we formulate the task of finding an optimal assembly plan as a search problem. In our model, a state S_n is formally defined by the set of completed edges. This set of edges represents the assembly's progress, conceptually tracking which components are fully assembled. The state transition actions are twofold: physical connection actions for actual assembly steps, and hierarchical navigation actions for shifts among sub-components.

However, a direct application of A* to this model presents two key challenges [17, 63]. First, if navigation is zero-cost, the A* evaluation function struggles to differentiate between an orderly plan and an inefficient one that chaotically switches between subtasks. Second, standard A* cannot determine the assembly order for independent subsystems that are disconnected in the graph.

To address these challenges, our A* algorithm optimizes not a single cost, but a multi-dimensional cost vector with prioritized components. This ensures that decisions prioritize physical progress before secondary goals like assembly coherence. We adapt the standard A* evaluation function's structure, but critically redefine its components:

$$f(S_n) = \overbrace{\sum_{a_i \in P(S_0,S_n)} c(a_i)}^{g(S_n):\text{Path Cost}} + \overbrace{\left(|\mathrm{R}(S_n)| + \sum_{g' \in \mathrm{U}(S_n)} |\mathrm{C}(g')| \right)}^{h(S_n):\text{Heuristic}}. \quad (1)$$

Although the number of remaining connections is countable in a certain graph, we frame $h(S_n)$ as a heuristic within the A* context to guide the search toward the goal state. The vector cost $g(S_n)$ and scalar heuristic $h(S_n)$ are combined through a prioritized tuple comparison logic.

In this formulation, the heuristic $h(S_n)$ is a scalar estimate of remaining physical connections, ensuring admissibility. Specifically, $\mathrm{R}(S_n)$ is the set of remaining connections in the current subtask, $\mathrm{U}(S_n)$ is the set of all unfinished subtasks, and $\mathrm{C}(g')$ is the set of connections within any given subtask g'.

By adapting the path cost $g(S_n)$ to a vector, the core change enables the A* search to evaluate paths using a sequence of objectives, first maximizing physical progress and subsequently penalizing incoherent switches between subtasks. To implement this, we define the action cost vector with components corresponding to physical connections, switches, and navigation steps, denoted as $c(a_i) = (c_{conn}, c_{switch}, c_{nav})$. A physical connection costs $(1,0,0)$, a coherent navigation is $(0,0,1)$, and an incoherent switch from an unfinished sub-assembly costs $(0,1,1)$. The total path cost $g(S_n)$ is the vector sum of these costs, yielding $g(S_n) = (g_{conn}, g_{switch}, g_{nav})$.

This vector directly forms the basis for a lexicographical comparison in the A* priority queue. Instead of a single value, the algorithm now evaluates states using a comprehensive sorting key:

$$SortKey = (g_{conn} + h,\ g_{switch},\ g_{nav},\ -depth). \quad (2)$$

This sorting key enforces a hierarchical comparison. The algorithm first compares the estimated physical progress to guarantee optimality, and only uses subsequent terms as tie-breakers to penalize incoherent paths.

To resolve component discreteness, we introduce a virtual root node connecting all independent sub-components. It transforms the inter-component sequence decision into a path-finding choice for the A* search. By optimizing the sorting key, the search yields a basic assembly path. This path with hierarchical coherence serves as the raw input for the subsequent LLM-mediated narration.

3.3 LLM-Mediated Narration Workflow for Generating Ergonomic Assembly Path

The assembly path produced by the A* search guarantees logical correctness but lacks ergonomic considerations for human operators. To transform this rigid sequence into intuitive instructions, we developed an agent workflow with three functional layers. This workflow separates the processing into data retrieval, ergonomic reasoning, and narration generation. We focus on the flow of assembly data between the different layers and state management to ensure that ergonomic refinements are grounded in the verifiable geometric and procedural facts of the assembly.

3.3.1 Assembly Path Structuring for Stable LLM Ingestion

The workflow begins with a structuring component, intended to translate the abstract assembly path into a structured, LLM-readable format. Establishing a stable and reliable data structure at the outset is a prerequisite for any downstream LLM-based processing. Through experimentation, we found that a direct, monolithic graph-to-JSON conversion was prone to structural errors. Consequently, we implemented a two-step process where the path is first converted into a series of Cypher queries before being serialized to a final JSON sequence (S_{raw}). This use of an intermediate, syntactically verifiable representation serves as a structuring gateway, ensuring that subsequent stages operate on a consistent and reliable data foundation.

3.3.2 Ergonomic Optimization of Assembly Path

The central ergonomic transformation is carried out by a set of LLM modules built on LangGraph. The **Optimizer** re-orders and groups steps in S_{raw} according to human-factors principles such as limiting tool switches and co-locating actions in the same physical zone. Task allocation and coordination are achieved through a strict separation of concerns: the **Optimizer** makes planning decisions while the **Retriever** supplies factual context on demand; the resulting structured steps are then handed to the **Narrator** in the final part for language refinement.

Engaging in ergonomic refinement requires access to information beyond the linear sequence of steps, including the physical properties of parts and their spatial relationships. The **Optimizer** therefore dynamically acquires this context by invoking the **Retriever** to query the assembly hypergraph for targeted facts. These queries might seek to identify the specific tool needed for a connection, determine the proximity of different components, or retrieve other relational data for applying a human-factors principle. To maintain focus and reduce informational noise, retrievals are constrained to the top-5 most relevant facts. This design grounds the **Optimizer**'s reasoning in the verifiable data of the hypergraph. Such an architectural separation of concerns makes the entire process more robust and transparent.

3.3.3 Narration Generation for Cognitive Fluency

The final stage of the workflow converts the optimized but abstract assembly path into an ergonomic guidance script for the AR environment. This task is performed by a dedicated **Narrator** module. This module operates in a two-tier process. A global planner first segments the task into coherent acts, then translates logical steps into specific UI commands. These commands will invoke a predefined set of narration tools within the AR client. To ensure consistency, the tool-calling is restricted to a strict schema of abstract actions. These actions are mapped to Unity prefabs and functions. Table 1 shows all narrative tools managed by the **Narrator**. Finally, a validation step verifies that the output sequence strictly aligns with the input assembly path.

Table 1: Mapping of Narrative Tools to AR Actions. The Narrator utilizes these tools as narrative primitives, which are finally rendered in Unity.

Tool	Inputs	AR Visualization
Text Hint	`text`	Displays a UI panel with text.
Highlight	`target_id`	Rotates model to face user and highlights part.
Link	`id_A`, `id_B`	Pulses a linking line between parts.
Hierarchy	`order_list`	Sequentially highlights all parts in a subgraph.
Animate	`id`, `edge`, `mode`	Plays install/remove animated instructions based on assembly data.
Check	`id_list`	Shows 3D prefabs of parts for user.

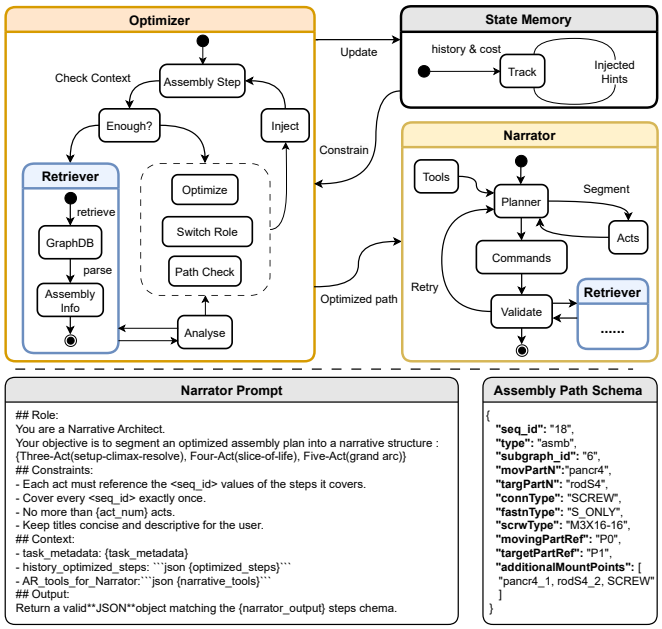

Fig. 3: **The implementation details of agent workflow.** The upper diagram illustrates the state transition logic where the Optimizer refines the path using retrieved context, and the Narrator structures the global narrative. The lower panels display the prompt for narrative planning and Data Schema used to ground the LLM.

Isolating this final translation step creates a separation between the two core agent layers. The Optimizer performs the planning by determining the content and sequence of tasks, whereas the Narrator handles the presentation by formulating the best method to convey those instructions to the user. This architectural separation enhances modularity, allowing the set of narration tools to be expanded or modified without impacting the upstream optimization logic. Thus, this orchestration of visual cues and texts transforms a single logical directive into a rich instruction.

3.3.4 Implementation Details

We implemented the agentic workflow using the LangGraph framework to manage state transitions and memory persistence. The system utilizes specific models from the Gemini family depending on the tasks. Figure 3 illustrates the detailed state transitions, sample prompts, and data schema. The **Optimizer** employs `gemini-2.5-pro` with a temperature of 0.5 to balance creative routing strategies with adherence to constraints. We allocated a thinking budget of 4000 tokens to support its reasoning over the optimization history. The **Narrator** also uses `gemini-2.5-pro` but operates with a temperature of 0.8 and an extended thinking budget of 8000 tokens to enhance narrative fluency and adaptability. In contrast, the **Retriever** utilizes the lightweight `gemini-2.5-flash` model with a temperature of 0.5 and a 2000 token budget for a low-latency retrieval when grounding information from the graph database.

4 User Study

To empirically evaluate our proposed LLM-mediated narration approach based on assembly hypergraph and understand the contributions of its representational and conveyance components, we conducted a within-subjects user study. This study was designed to compare our system against two key baselines: a hypergraph-only system that exposes structural information without LLM-mediated narration, and a traditional liaison graph system representing the conventional approach to AR guidance.

4.1 Research Questions and Hypotheses

Our study was guided by two primary research questions that examine the sequential contributions of structured representation and LLM-mediated narration workflow.

RQ1: Compared to a conventional liaison graph, does a hypergraph representation reduce the procedural difficulty of a complex assembly task and improve objective outcomes?

H1.1 Cognitive Process: Compared to the liaison graph, the structured guidance of the hypergraph will reduce users' overall cognitive load. In terms of attention patterns, it will enhance operational efficiency in the physical assembly area, though this structural information may come at the cost of increased attentional switching frequency.

H1.2 Task Outcome: These process-level improvements will result in task completion times (TTC) that are no worse than the liaison graph condition and error rates that are equal to or lower.

H1.3 Attentional Cost: The structural advantages offered by the Hypergraph in deconstructing complex tasks come at the cost of increased cognitive switching. This will be observed as more frequent attentional shifts of gaze events between different information areas compared to the liaison graph.

RQ2: Building on the Hypergraph condition, does introducing the LLM-Mediated Narration further optimize cognitive burden and subjective evaluations while maintaining or enhancing objective performance?

H2.1 Attention Guidance: The LLM-Mediated Narration method will further optimize users' attentional allocation. Specifically, the narration instructions will significantly reduce reliance on low-information-density animated models, redirecting attention to the higher information density text instructions while simultaneously reducing the cognitive switching cost in the task area.

H2.2 Overall Evaluation: These experiential enhancements will be reflected in higher ratings for immersion (ARI), usability (SUS), and technology acceptance (TAM). Concurrently, objective performance metrics such as task completion time (TTC) and error rates will be at least as good as, if not better than, the Hypergraph-only condition.

4.2 Study Design

4.2.1 Task and Conditions

Participants were tasked with assembling a multi-part robotic gripper, a task representative of moderately complex industrial procedures involving 25 components and 20 pairs of fasteners. The experiment employed a within-subjects design where each participant completed the assembly task under all three conditions: **(1) liaison graph:** A baseline method that presented a linear sequence of steps derived from a simple liaison graph. It offers standard context, showing only the immediate next connection without grouping related sub-tasks. **(2) Hypergraph:** An experimental condition using our hypergraph representation to provide structured, stage-aware guidance. While it groups steps by sub-assembly stages, it presents this structure in a raw format without narrative optimization. **(3) LLM+HG:** Our full proposed method, which used the LLM-mediated narration workflow to translate the structure of the hypergraph into an ergonomically optimized and coherent sequence of instructions, explicitly managing attention shifts and summarizing complex actions. The order of conditions was counterbalanced across participants using a Latin square design to mitigate learning and fatigue effects.

4.2.2 Participants and Procedure

We recruited 24 participants (17 male, 7 female) with engineering backgrounds from a local university ($M = 23.75, SD = 2.63$). The Northwest University Medical Ethics Committee approved this study (IRB 241113089, Nov 13, 2024), and all participants signed written informed consent. Given that most participants had limited prior AR experience, a standardized tutorial session was conducted to ensure basic HMD proficiency. Figure 4 illustrates the specific procedure of the study. Participants performed the assembly task under three counterbalanced conditions. Mandatory rest periods were enforced between trials to prevent fatigue. Subjective questionnaires were completed after each condition, concluding with a semi-structured interview.

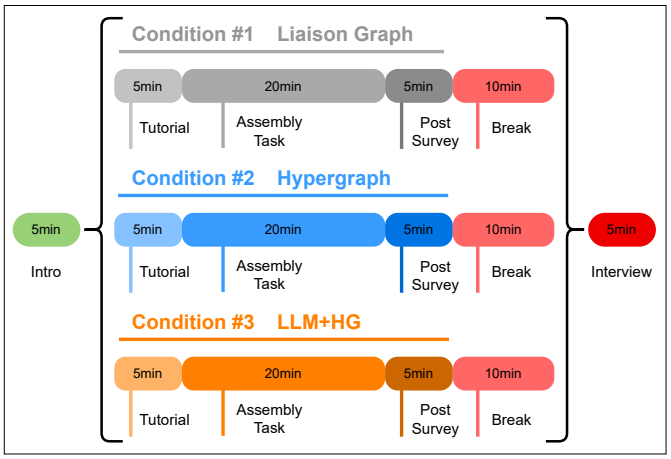

Fig. 4: **The timeline for the within-subjects study.** Participants engaged with three conditions in counterbalanced sessions. Each session included a tutorial, the core task of AR assembly, and a post-task survey, interspersed with breaks and concluding with a semi-structured interview.

4.2.3 Apparatus and Wizard-of-Oz

The study was conducted using a Pico 4 Ultra HMD (90Hz refresh rate, 104° FoV) with video-see-through capabilities, connected via Wi-Fi to a workstation. To simulate robust state recognition, we employed a Wizard-of-Oz (WoZ) methodology [14,54]. The operator controlled the system via a desktop GUI comprising three information streams: (1) a real-time first person view stream cast from the participant's HMD to verify fine manipulations; (2) a direct view of the physical workspace to monitor actions potentially occluded from the HMD cameras; (3) a logic panel displaying the current step index and state constraints.

A single trained experimenter conducted all sessions. The average system response latency including operator reaction time was recorded at approximately 250 ms. Participants were unaware of human interventions and were led to believe that the system was fully automated. This design was chosen to ensure internal validity by eliminating tracking latency and recognition failures as confounding variables, thereby isolating the effects of the guidance method.

4.3 Measurements

We quantified participant performance and experience using a combination of objective and subjective metrics. Objective performance was primarily assessed by Time To Completion (TTC), measured from the start to the completion of the assembly, and Error Rates, which included both total errors and specific procedural failures. To analyze cognitive effort and attention allocation, we captured head gaze data, from which we calculated the normalized dwell duration and transition counts within three predefined Areas of Interest (AOI): the Production Model Area, the Assembly Area, and the Instruction Area.

Following each condition, participants' subjective experiences were assessed. Perceived task load was measured using the NASA Task Load Index (NASA-TLX). System usability was evaluated with the System Usability Scale (SUS), immersion with the Augmented Reality Immersion (ARI) scale, and user acceptance via the Technology Acceptance Model (TAM). After the final condition, we conducted a semi-structured interview focused on participants' perceptions of the differences between the guidance systems and how these differences impacted their problem-solving approaches

5 Results

We present the findings from the user study, aimed at quantifying the distinct contributions of the Hypergraph representation and the LLM-mediated narration. All statistical analyses were prespecified to examine the effects of the different guidance modalities on objective performance, attention patterns, and subjective experience. For our repeated measures analysis of variance, we first conducted Mauchly's test of sphericity [42]. When this assumption was violated ($p < .05$), we report Greenhouse-Geisser corrected results [21]. All post hoc pairwise tests were adjusted for multiple comparisons using the Holm–Bonferroni method [28]. Effect sizes are reported using Cohen's d, for which 95% confidence intervals (CIs) are also provided [13]. For measures where a lower score indicates better performance, a negative effect size signifies an improvement for the intervention group over the control group.

Table 2: Time To Completion (TTC) in seconds. Statistical metrics (p, Cohen's d) represent pairwise comparisons against the Liaison graph baseline.

Method	Avg. ±std.dev.	p	Cohen's d	Effect size
Liaison graph	992.32 ± 237.91	-	-	-
Hypergraph	873.6 ± 117.51	0.012	0.632	Medium
LLM+HG	826.50 ± 146.2	0.047	0.839	Large

5.1 Validation of Experimental Design

To verify whether the presentation order introduced confounding factors to our within-subjects design, we conducted RM-ANOVAs with presentation order (1st, 2nd, 3rd) as the independent variable. No significant main effects were found for Time To Completion (TTC) (F(2, 46) = 1.19, p = .32), Total Error Rate (F(2, 46) = 1.05, p = .36), or NASA-TLX (F(2, 46) = 0.96, p = .40). Interestingly, TTC trended upward ($M_{1st} = 836s$ vs. $M_{3rd} = 940s$), suggesting physical fatigue from prolonged AR usage likely counteracted typical learning curves. While errors decreased slightly ($M_{1st} = 0.136$ vs. $M_{3rd} = 0.098$), the lack of statistical significance confirms that our Latin square design effectively distributed these opposing factors (fatigue vs. learning). Consequently, the results reported below can be attributed to the guidance modalities.

5.2 Objective Results

5.2.1 Performance Outcomes

We analyzed Time To Completion (TTC) and error rates to assess operational efficiency. The mean task completion times are shown in Table 2. The main effect from the RM-ANOVA on TTC was statistically significant (F(2, 46) = 3.32, p = .045). However, with the CIs upper bound nearing zero, the significance was not robust enough (d = -0.84, p = .047, 95% CI [-1.55, -0.02]). The full LLM-mediated narration method showed a large improvement over the liaison graph baseline. Although the effect reached statistical significance, its marginal nature warrants cautious interpretation.

In contrast, while pairwise p-values for the other comparisons were low, their CIs for the effect sizes both crossed zero (d = -0.63, 95% CI [-0.93, 0.05], d = -0.36, 95% CI [-0.93, 0.28]). This suggests that while trends for improvement exist, these effects were not strong enough to be considered statistically significant.

A similar pattern was observed in error rates, as illustrated in Figure 5. Although the overall RM-ANOVA for Total Error Rate did not reach statistical significance (F(2, 46) = 2.08, p = .134), likely due to high variance across conditions, planned pairwise comparisons revealed an advantage. Specifically, the full method led to a significant reduction in total errors compared to the liaison graph baseline (p = .025, d = -0.52), highlighting a targeted improvement in accuracy not captured by the omnibus test. A deeper analysis into error subtypes showed that the ANOVA for Assembly Failure Rate was significant (F(2, 46) = 4.75, p = .013). The comparison between the LLM-mediated narration and hypergraph conditions on this metric is particularly revealing. While its post-hoc p-value was marginal (p = .051), the confidence interval for the effect size (d = -0.74, 95% CI [-1.41, 0.02]) indicated a strong trend towards a reduction in errors, though the interval's upper bound slightly crossed zero. This suggests that while not definitively confirmed with 95% confidence, there is evidence for the LLM-mediated narration method's superiority in reducing assembly failures over the hypergraph. In contrast, for other error-rate comparisons with low p-values, the CIs more clearly contained zero, indicating these effects

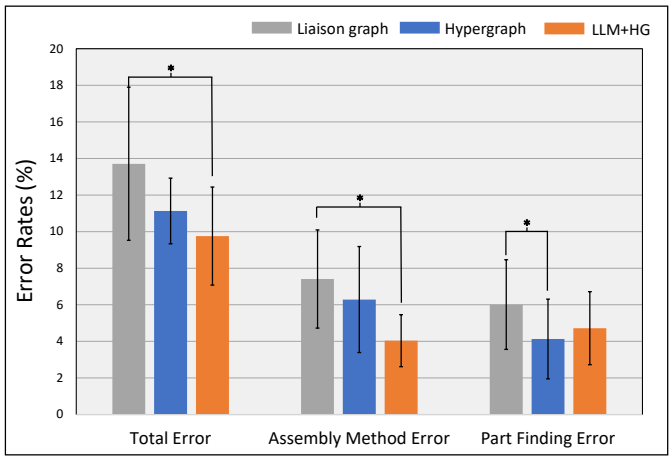

Fig. 5: **Comparison of error rates across conditions.** The chart displays Total Error rates alongside a breakdown into Assembly Method Error and Part Finding Error. Error bars indicate standard error. Asterisks (*) denote pairwise significance ($p < .05$).

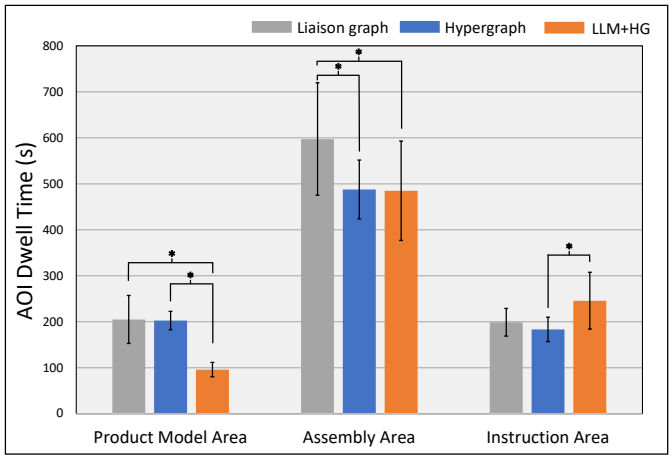

Fig. 6: **Fixation dwell time on Areas of Interest (AOI).** Attention patterns are quantified by the total duration of fixations in each area. Error bars indicate standard error. Asterisks (*) denote pairwise significance ($p < .05$).

were not robust. This underscores that the LLM-mediated narration's primary benefit was in clarifying how the assembly process should be performed, thereby preventing procedural missteps.

5.2.2 Attention Patterns

To understand the cognitive processes underlying the performance outcomes, we analyzed head gaze data, focusing on how participants allocated their attention across three predefined Areas of Interest. To isolate the effects of the guidance modality from variations in task duration, the following analysis reports on the normalized duration and counts.

Figure 6 visualizes the total dwell time within each AOI across conditions. A significant main effect was found for dwell time ($F(2, 46) = 42.18$, $p < .001$). Post hoc tests revealed that the LLM-mediated narration method reduced reliance on animated demonstrations, with participants spending significantly less time viewing the model than in both the liaison graph (d = -1.73, 95% CI [-2.81, -0.94]) and hypergraph (d = -1.26, 95% CI [-1.62, -0.53]) conditions. This very large effect suggests that the LLM-mediated narration's instructions were sufficiently clear and information-dense to obviate the need for frequent visual reference to the animated model. Both the hypergraph and LLM-mediated narration conditions also led to less dwell time in the

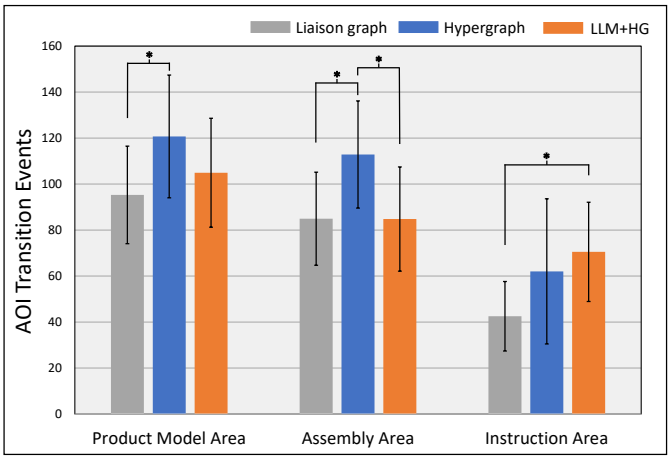

Fig. 7: **Transition events on Areas of Interest (AOI).** Attention patterns are quantified by the frequency of transition events. Error bars indicate standard error. Asterisks (*) denote pairwise significance ($p < .05$).

physical Assembly Area compared to the baseline, indicating higher efficiency. Conversely, gaze duration on the text instructions was highest in the LLM-mediated narration condition, reflecting the increased informational content provided by its human-factors-informed textual prompts.

Analysis of transition events, shown in Figure 7, further illuminates the cognitive cost of information processing. The ANOVA for transition events in the Assembly Area was significant ($F(2, 46) = 17.08$, $p < .001$). The results indicate that the benefits of the raw hypergraph structure came at the cost of increased attentional switching, as participants in this condition gazed more frequently at the Assembly Area compared to the LLM-mediated narration method (d = -1.22, 95% CI [-1.58, -0.94]). The LLM-mediated narration method successfully mitigated this cognitive overhead, reducing the switching frequency back to a level statistically indistinguishable from the baseline (p = .13). This finding suggests that while a structured representation is beneficial, an LLM-mediated narration is critical for integrating that structure in a way that minimizes disruptive and costly context switching.

5.3 Subjective Questionnaire Results

5.3.1 System Usability Scale (SUS)

To assess perceived usability, we utilized the SUS, a widely recognized and reliable tool for measuring the subjective usability of a system [8]. The SUS scores depicted in Figure 8 (a) demonstrate that the structural and narration enhancements had independent and cumulative positive effects on perceived usability. The RM-ANOVA was highly significant ($F(2, 46) = 42.20$, $p < .001$), and crucially, all pairwise comparisons between the three conditions were also significant, with 95% CIs for the effect sizes robustly excluding zero. This clear, stepwise improvement underscores that both the underlying representational model and the method of its conveyance are critical, interacting components of a usable and effective AR guidance method.

5.3.2 NASA Task Load Index (NASA-TLX)

The NASA-TLX is a standard multidimensional assessment of subjective cognitive workload [25]. We utilized it to quantify the cognitive burden on participants under each guidance condition. The results, shown in Figure 8 (b) corroborate the patterns observed in the objective data. The RM-ANOVA revealed a significant main effect of guidance condition on perceived task load ($F(2, 46) = 12.11$, $p < .001$). The introduction of the hypergraph structure was a key factor in reducing cognitive burden, leading to significantly lower TLX scores compared to the liaison graph baseline (d = -0.77, 95% CI [-1.41, -0.39]). The LLM-mediated narration method further reduced the mean workload score, resulting in a large and significant improvement over the baseline (d = -0.84, 95% CI [-1.52, -0.79]). While the numerical difference

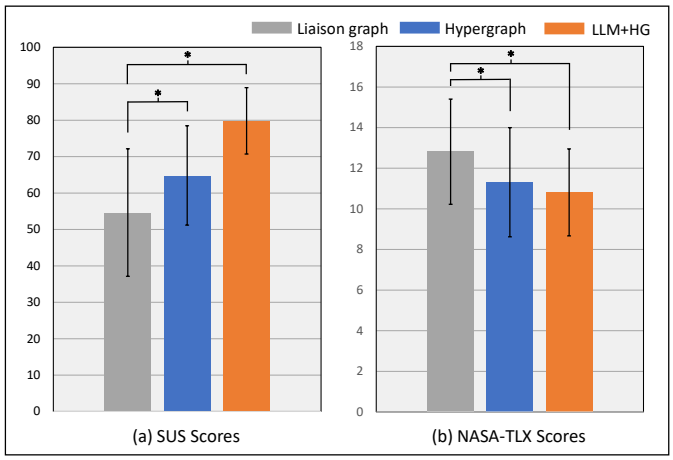

Fig. 8: **Assessment of usability (SUS) and cognitive workload (NASA-TLX).** Higher SUS scores indicate better usability, while lower NASA-TLX scores indicate reduced workload. Error bars indicate standard error. Asterisks (*) denote pairwise significance ($p < .05$).

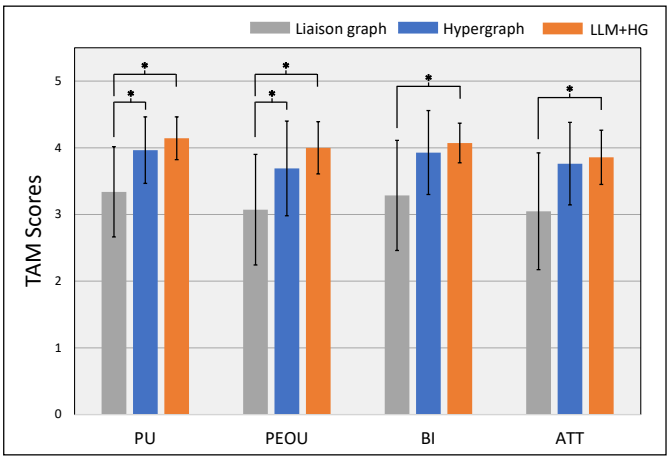

Fig. 9: **Technology Acceptance Model (TAM) scores.** The evaluation covers four dimensions: Perceived Usefulness (PU), Perceived Ease of Use (PEOU), Behavioral Intention (BI), and Attitude (ATT). Error bars indicate standard error. Asterisks (*) denote pairwise significance ($p < .05$).

between the LLM-mediated narration and hypergraph conditions did not reach statistical significance, as its CI crossed zero, the overall trend suggests that the structural representation provided by the hypergraph addressed the majority of the cognitive strain.

5.3.3 Technology Acceptance Model (TAM)

The Technology Acceptance Model (TAM) is a foundational theory explaining how users accept new technology based on its Perceived Usefulness (PU), Perceived Ease of Use (PEOU), Behavioral Intention (BI), and Attitude (ATT) [15]. We applied this model to evaluate participants' acceptance of each guidance method. The results, shown in Figure 9, reveal that acceptance was primarily driven by the structural improvements of the hypergraph. For the core dimensions of PU, BI, and ATT, both the hypergraph and LLM-mediated narration conditions significantly outperformed the liaison graph baseline, but were not significantly different from each other, as confirmed by 95% confidence intervals that contained zero. The LLM-mediated narration's contribution was more nuanced, manifesting in the PEOU dimension. Here, while a positive trend was observed over the hypergraph condition, the difference was not statistically significant as the CI also crossed zero (95% CI [-0.09, 2.43]), suggesting the full method's refined interaction enhanced the quality of the experience, though not to a degree that was robustly distinguishable from the hypergraph alone.

5.3.4 Augmented Reality Immersion (ARI)

The Augmented Reality Immersion (ARI) questionnaire is a validated scale for measuring user immersion in AR environments [20]. We used it to assess the multi-faceted nature of immersion generated by each method, with the results shown in Figure 10. The analysis revealed that the LLM-mediated narration method provided a richer and more comprehensively immersive experience. Greenhouse-Geisser corrections were applied to the ANOVA results for the Interest, Usability, and Presence dimensions, as Mauchly's test indicated a violation of the sphericity assumption. The LLM-mediated narration condition was rated significantly higher than the liaison graph baseline on Usability (d = 1.52, 95% CI [0.79, 2.14]), Interest (d = 0.93, 95% CI [0.62, 1.66]), and Presence (d = 0.93, 95% CI [0.25, 1.60]). In contrast, the hypergraph only significantly improved upon Interest and Usability. Notably, the Usability dimension exhibited a clear stepwise progression, with the LLM-mediated narration being superior to the hypergraph (d = 0.54, 95% CI [0.21, 1.26]), which was in turn superior to the liaison graph (d = 0.51, 95% CI [0.02, 1.98]), mirroring the SUS results.

Furthermore, confidence intervals provided stronger evidence for the LLM-mediated narration's ability to induce deeper immersion than p-values alone. For the comparison with the hypergraph, the 95%

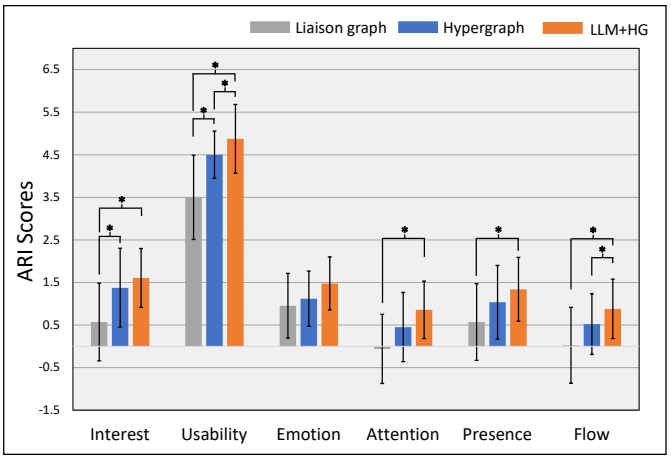

Fig. 10: **Augmented Reality Immersion (ARI) scores.** The questionnaire measures user immersion across dimensions including Interest, Usability, Emotion, Attention, Presence, and Flow. Error bars indicate standard error. Asterisks (*) denote pairwise significance ($p < .05$).

CI for the effect on Flow was entirely above zero (d = 1.07, 95% CI [0.18, 1.70]), indicating a significant positive effect. For Focus of Attention, the CI excluded zero (d = 1.00, 95% CI [0.02, 1.11]), providing strong evidence for a positive trend. These findings indicate that while a structured hypergraph can make a task more interesting, the ergonomic optimizations of LLM-mediated narration method are crucial for elevating the experience into one that is usable, engaging, and creates a stronger sense of presence and flow.

6 Discussion

The user study findings indicate a progressive enhancement in assembly guidance effectiveness, beginning with the structural foundation of the hypergraph and culminating in the refinements of the LLM-mediated narration. This progression was observable across objective performance, attentional patterns, and subjective experiences. The choice of guidance modality influenced task completion time, with a discernible tendency towards greater efficiency, particularly when the full method was compared to the liaison graph baseline. This performance trend was accompanied by a notable shift in attentional strategy. The guidance appeared to reduce reliance on animated models in favor of more information-dense text instructions. Furthermore, while

the introduction of the hypergraph structure was associated with increased attentional switching in the assembly area, the LLM-mediated narration workflow seemed to counteract this effect, bringing AOI transition counts closer to baseline levels, optimizing visual search efficiency [1, 52]. This pattern aligns with subjective reports, which showed a marked reduction in cognitive load and a clear, stepwise improvement in system usability.

The initial improvements in performance and usability appear to stem from the representational capacity of the hypergraph. By encoding prerequisite relationships and a hierarchically consistent workflow, the hypergraph model constrains the set of possible actions in a way comparable to hierarchical task planning [2, 4, 11, 16], thereby potentially reducing procedural errors that necessitate rework. This structural enforcement may help explain the observed decrease in dwell time within the physical assembly area and the reported reduction in cognitive load, as participants were guided along a validated, logical path without split-attention redundancy [9, 46]. These benefits are reflected in the improved ratings as enhanced user acceptance. However, this added clarity was not without an apparent cost. As hypothesized (H1.3), the explicit presentation of sub-assemblies and stage transitions was associated with more frequent attentional shifts in the hypergraph-only condition, suggesting a cognitive overhead involved in mapping the abstract structure to the physical workspace. This observation is consistent with related work [30], suggesting that while exposing users to a task's causal hierarchy can improve objective task outcomes, it may also impose a higher initial processing cost, highlighting a tension between representational expressiveness and immediate cognitive accessibility.

The LLM-mediated narration layer appeared to function as a mediating component, reconfiguring the cognitive and attentional dynamics. Its primary contribution seems to be the semantic reorganization of the algorithm-optimal path into ergonomically coherent chunks. By clustering steps based on tool usage and component locality, the LLM-mediated narration was designed to minimize the need for disruptive context switches, which ground instructions in physical affordances [7, 11]. This is supported by the reduction in transition counts within the assembly area from their peak in the hypergraph condition. The full method's narratization also altered how information was consumed. The dense, action-oriented textual instructions offered a more efficient pathway for information acquisition compared to the animated models. This reallocation of attention should therefore be interpreted not as a transfer of cognitive burden, but as a shift to a more direct mode of instruction. This, in turn, was accompanied by the stepwise enhancements in usability and dimensions of immersion, particularly Presence and Flow, suggesting the narration layer is an important intermediary for translating a structurally sound plan into a human-executable process.

To summarize, the findings provide nuanced answers to our research questions. Regarding RQ1, the hypergraph representation was associated with a reduced perceived task difficulty (H1.1), and while objective performance metrics showed a favorable trend, the statistical evidence was not universally conclusive after correction (H1.2). The anticipated attentional cost of this structural advantage was observed, supporting H1.3. In response to RQ2, the LLM-mediated narration optimized attentional processes by guiding focus and alleviating the switching burden introduced by the raw hypergraph (H2.1). These process-level refinements translated into higher usability and immersion ratings, while objective performance was maintained (H2.2). The evidence from this study suggests a complementary relationship between the two core components of our method. The hypergraph representation appears to offer a structural foundation that ensures procedural correctness and reduces baseline cognitive load. The LLM-mediated narration offers potential ergonomic and experiential refinements that optimize the flow of interaction.

7 Limitations and Future Work

In this section, we discuss several limitations of our findings within the context of our study and outline potential directions for future research.

Validation in Real-world Industrial Settings. While prior research has often utilized simplified assembly tasks to minimize experimental variables, our study targeted a complex assembly to induce cognitive load, which is often underrepresented in simplified setups. However, a limitation lies in our participant pool, which consisted primarily of engineering students. In real-world industrial contexts, skilled operators often possess rich knowledge and may exhibit different cognitive patterns compared to novices. As a result, the generalizability of our findings to expert workers remains to be verified. Future work will involve deploying the system in actual production environments and recruiting experienced technicians to investigate how the system adapts to users with varying levels of expertise.

Real-Time Adaptive Guidance. In the current implementation, the LLM narration functions primarily as an offline content optimizer validated via a Wizard-of-Oz experiment. This design was used to ensure the internal validity of the experiment, allowing us to isolate the effects of the narrative strategy from system-level artifacts such as perception jitter or sensing latency. However, we acknowledge that this limits the system's ability to dynamically adapt to unexpected user behavior or error recovery in real-time. Future work aims to evolve this framework into a full system. We plan to integrate a vision-based 6D pose tracking module grounded in the assembly hypergraph to enable real-time state awareness. Furthermore, recognizing the challenges of edge computing, we will investigate guidance strategies to ensure stable instruction delivery during high-load inference periods.

Scalability to Complex Scenarios. Another area worth exploring is the system's scalability across other industrial operations. First, regarding experimental design, the current study relied on a single assembly task. Although we employed a counterbalanced design to mitigate ordering effects and statistical analysis showed no significant order effects, potential residual learning effects and the specific nature of the chosen task may limit the immediate generalizability of our findings. Future investigations will prioritize diverse assembly scenarios to validate robustness. Second, from a technical perspective, the current framework effectively optimizes discrete rigid body kinematics but faces structural limitations when scaling to spatially continuous operations, such as electrical routing or fluid application, as noted in the constraints of our A* cost function. We envision evolving the data structure towards a geometry-centric hypergraph grounded with the 3D model, thereby enabling the path planning algorithm to support mixed-domain constraints and non-rigid assembly procedures.

8 Conclusion

In this work, we proposed a novel AR assembly guidance framework that synergizes a semantically enriched hypergraph with an LLM-mediated narration workflow. By formalizing assembly tasks into hypergraphs, our method explicitly captures complex procedural constraints often missed by conventional liaison graphs. We further employed a structured LLM agent pipeline to translate rigid assembly paths into ergonomically fluent narratives, effectively bridging the gap between algorithmic logic and human cognitive needs. An IRB-approved within-subjects user study demonstrated that this approach was associated with significantly reduced cognitive load and enhanced usability and immersion compared to baselines, while maintaining high task performance. These findings support the potential of using LLMs as a semantic mediation layer for complex AR instructions. Future work will focus on building real-time systems, and extending the framework to contribute towards a standardized engineering language for AR guidance.

Acknowledgments

This work is partially supported by National Natural Science Foundation of China through Project (62571051), by Open Project Program of State Key Laboratory of Virtual Reality Technology and Systems, Beihang University (No. VRLAB2024C02), by General Projects of the Shaanxi Provincial Department of Science and Technology (2025JC-YBQN-801).

REFERENCES

[1] I. B. Adhanom, P. MacNeilage, and E. Folmer. Eye Tracking in Virtual Reality: a Broad Review of Applications and Challenges. *Virtual Reality*, 27(2):1481–1505, June 2023. doi: 10.1007/s10055-022-00738-z 2, 9

[2] C. Aeronautiques, A. Howe, C. Knoblock, I. D. McDermott, A. Ram, M. Veloso, D. Weld, D. W. Sri, A. Barrett, D. Christianson, and others. Pddl—the planning domain definition language. *Technical Report, Tech. Rep.*, 1998. 2, 9

[3] 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*, 23(18):7698, Sept. 2023. doi: 10.3390/s23187698 2

[4] R. Aoki, Y. Cao, S. Mahdavi, and K. Tang. Leveraging Environment Interaction for Automated PDDL Translation and Planning with Large Language Models. In *Advances in Neural Information Processing Systems 37*, pp. 38960–39008. Neural Information Processing Systems Foundation, Inc. (NeurIPS), Vancouver, BC, Canada, 2024. doi: 10.52202/079017-1230 2, 9

[5] L. Becker, T. Nilsson, P. Demedeiros, and F. Rometsch. Augmented Reality in Service of Human Operations on the Moon: Insights from a Virtual Testbed. In *Extended Abstracts of the 2023 CHI Conference on Human Factors in Computing Systems*, pp. 1–8. ACM, Hamburg Germany, Apr. 2023. doi: 10.1145/3544549.3585860 2

[6] V. Biener, S. Kalamkar, N. Nouri, E. Ofek, M. Pahud, J. J. Dudley, J. Hu, P. O. Kristensson, M. Weerasinghe, K. C. Pucihar, M. Kljun, S. Streuber, and J. Grubert. Quantifying the Effects of Working in VR for One Week. *IEEE Trans. Visual. Comput. Graphics*, 28(11):3810–3820, Nov. 2022. doi: 10.1109/TVCG.2022.3203103 2

[7] A. Brohan, Y. Chebotar, C. Finn, K. Hausman, A. Herzog, D. Ho, J. Ibarz, A. Irpan, E. Jang, R. Julian, and others. Do as i can, not as i say: Grounding language in robotic affordances. In *Conference on robot learning*, pp. 287–318. PMLR, 2023. 2, 9

[8] J. Brooke et al. SUS: A 'Quick and Dirty' Usability Scale. In P. W. Jordan, B. Thomas, I. L. McClelland, and B. Weerdmeester, eds., *Usability Evaluation In Industry*, pp. 207–212. CRC Press, 0 ed., June 1996. doi: 10.1201/9781498710411-35 7

[9] S. Cammeraat, G. Rop, and B. B. De Koning. The influence of spatial distance and signaling on the split-attention effect. *Computers in Human Behavior*, 105:106203, Apr. 2020. doi: 10.1016/j.chb.2019.106203 2, 9

[10] P. Carlson, A. Peters, S. B. Gilbert, J. M. Vance, and A. Luse. Virtual Training: Learning Transfer of Assembly Tasks. *IEEE Trans. Visual. Comput. Graphics*, 21(6):770–782, June 2015. doi: 10.1109/TVCG.2015.2393871 1

[11] G. Chen, L. Yang, R. Jia, Z. Hu, Y. Chen, W. Zhang, W. Wang, and J. Pan. Language-Augmented Symbolic Planner for Open-World Task Planning. In *Robotics: Science and Systems XX*. Robotics: Science and Systems Foundation, July 2024. doi: 10.15607/RSS.2024.XX.037 2, 9

[12] M. Chen, M. Peljhan, and M. Sra. ConnectVR: A Trigger-Action Interface for Creating Agent-based Interactive VR Stories. In *2024 IEEE Conference Virtual Reality and 3D User Interfaces (VR)*, pp. 286–297. IEEE, Orlando, FL, USA, Mar. 2024. doi: 10.1109/VR58804.2024.00051 2

[13] J. Cohen. *Statistical Power Analysis for the Behavioral Sciences*. Routledge, 0 ed., May 2013. doi: 10.4324/9780203771587 6

[14] N. Dahlbäck, A. Jönsson, and L. Ahrenberg. Wizard of Oz studies — why and how. *Knowledge-Based Systems*, 6(4):258–266, Dec. 1993. doi: 10.1016/0950-7051(93)90017-N 6

[15] F. D. Davis. Perceived Usefulness, Perceived Ease of Use, and User Acceptance of Information Technology. *MIS Quarterly*, 13(3):319–340, Sept. 1989. doi: 10.2307/249008 8

[16] K. Erol, J. A. Hendler, and D. S. Nau. UMCP: A sound and complete procedure for hierarchical task-network planning. In *Aips*, vol. 94, pp. 249–254, 1994. 2, 9

[17] B. Fu, L. Chen, Y. Zhou, D. Zheng, Z. Wei, J. Dai, and H. Pan. An improved A* algorithm for the industrial robot path planning with high success rate and short length. *Robotics and Autonomous Systems*, 106:26–37, Aug. 2018. doi: 10.1016/j.robot.2018.04.007 4

[18] H. Gao and E. Kasneci. Exploring Eye Tracking as a Measure for Cognitive Load Detection in VR Locomotion. In *Proceedings of the 2024 Symposium on Eye Tracking Research and Applications*, pp. 1–3. ACM, Glasgow United Kingdom, June 2024. doi: 10.1145/3649902.3655644 2

[19] M. Gattullo, A. Evangelista, A. E. Uva, M. Fiorentino, and J. L. Gabbard. What, How, and Why are Visual Assets Used in Industrial Augmented Reality? A Systematic Review and Classification in Maintenance, Assembly, and Training (From 1997 to 2019). *IEEE Trans. Visual. Comput. Graphics*, 28(2):1443–1456, Feb. 2022. doi: 10.1109/TVCG.2020.3014614 1, 2

[20] Y. Georgiou and E. A. Kyza. The development and validation of the ARI questionnaire: An instrument for measuring immersion in location-based augmented reality settings. *International Journal of Human-Computer Studies*, 98:24–37, Feb. 2017. doi: 10.1016/j.ijhcs.2016.09.014 8

[21] S. W. Greenhouse and S. Geisser. On Methods in the Analysis of Profile Data. *Psychometrika*, 24(2):95–112, June 1959. doi: 10.1007/BF02289823 6

[22] I. Grobelna and A. Karatkevich. Challenges in Application of Petri Nets in Manufacturing Systems. *Electronics*, 10(18):2305, Sept. 2021. doi: 10.3390/electronics10182305 2

[23] K. Han, Y. Wang, J. Guo, Y. Tang, and E. Wu. Vision GNN: an image is worth graph of nodes. In *Proceedings of the 36th International Conference on Neural Information Processing Systems*, NIPS '22. Curran Associates Inc., Red Hook, NY, USA, 2022. event-place: New Orleans, LA, USA. 2

[24] M. Han, Y. Zhu, S.-C. Zhu, Y. Wu, and Y. Zhu. INTERPRET: Interactive Predicate Learning from Language Feedback for Generalizable Task Planning. In *Robotics: Science and Systems XX*. Robotics: Science and Systems Foundation, July 2024. doi: 10.15607/RSS.2024.XX.034 2

[25] S. G. Hart. Nasa-Task Load Index (NASA-TLX); 20 Years Later. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 50(9):904–908, Oct. 2006. doi: 10.1177/154193120605000909 7

[26] S. J. Henderson and S. K. Feiner. Augmented reality in the psychomotor phase of a procedural task. In *2011 10th IEEE International Symposium on Mixed and Augmented Reality*, pp. 191–200. IEEE, Basel, Oct. 2011. doi: 10.1109/ISMAR.2011.6092386 1

[27] T. Hoang, S. Greuter, and S. Taylor. An Evaluation of Virtual Reality Maintenance Training for Industrial Hydraulic Machines. In *2022 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, pp. 573–581. IEEE, Christchurch, New Zealand, Mar. 2022. doi: 10.1109/VR51125.2022.00077 2

[28] S. Holm. A simple sequentially rejective multiple test procedure. *Scandinavian journal of statistics*, pp. 65–70, 1979. Publisher: JSTOR. 6

[29] A. Jacovi, A. Wang, C. Alberti, C. Tao, J. Lipovetz, K. Olszewska, L. Haas, M. Liu, N. Keating, A. Bloniarz, and others. The FACTS Grounding Leaderboard: Benchmarking LLMs' Ability to Ground Responses to Long-Form Input. *arXiv preprint arXiv:2501.03200*, 2025. 2

[30] R. Jain, J. Shi, A. Benton, M. Rasheed, H. Doh, S. Chidambaram, and K. Ramani. Visualizing Causality in Mixed Reality for Manual Task Learning: A Study. *IEEE Trans. Visual. Comput. Graphics*, 31(12):10200–10214, Dec. 2025. doi: 10.1109/TVCG.2025.3542949 2, 9

[31] C. Jiang, B. Qi, X. Hong, D. Fu, Y. Cheng, F. Meng, M. Yu, B. Zhou, and J. Zhou. On Large Language Models' Hallucination with Regard to Known Facts. In *Proceedings of the 2024 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (Volume 1: Long Papers)*, pp. 1041–1053. Association for Computational Linguistics, Mexico City, Mexico, 2024. doi: 10.18653/v1/2024.naacl-long.60 2

[32] K. Kim, M. Billinghurst, G. Bruder, H. B.-L. Duh, and G. F. Welch. Revisiting Trends in Augmented Reality Research: A Review of the 2nd Decade of ISMAR (2008–2017). *IEEE Trans. Visual. Comput. Graphics*, 24(11):2947–2962, Nov. 2018. doi: 10.1109/TVCG.2018.2868591 1

[33] S. Kim, D. Kim, J.-E. Shin, and W. Woo. Object Cluster Registration of Dissimilar Rooms Using Geometric Spatial Affordance Graph to Generate Shared Virtual Spaces. In *2024 IEEE Conference Virtual Reality and 3D User Interfaces (VR)*, pp. 796–805. IEEE, Orlando, FL, USA, Mar. 2024. doi: 10.1109/VR58804.2024.00099 2

[34] P. Kourtesis, R. Amir, J. Linnell, F. Argelaguet, and S. E. MacPherson. Cybersickness, Cognition, & Motor Skills: The Effects of Music, Gender, and Gaming Experience. *IEEE Trans. Visual. Comput. Graphics*, 29(5):2326–2336, May 2023. doi: 10.1109/TVCG.2023.3247062 2

[35] G. Laouénan, J.-Y. Didier, and P.-E. Dossou. Performance and ergonomics of automated versus manual validation for AR-supervised industrial operations. In *2025 IEEE Conference Virtual Reality and 3D User Interfaces (VR)*, pp. 135–145. IEEE, Saint Malo, France, Mar. 2025. doi: 10.1109/VR59515.2025.00038 2

[36] E. Laviola, M. Gattullo, S. Romano, and A. E. Uva. Which Side is the Top? A User Study to Compare Visual Assets for Component Orientation in Assembly with Augmented Reality. *IEEE Trans. Visual. Comput. Graphics*, 31(5):3470–3480, May 2025. doi: 10.1109/TVCG.2025.3549164 2

[37] L. S. Lenz, A. R. Fender, J. Chatain, and C. Holz. Comparing Synchronous and Asynchronous Task Delivery in Mixed Reality Environments. *IEEE Trans. Visual. Comput. Graphics*, 30(5):2776–2784, May 2024. doi: 10.1109/TVCG.2024.3372034 2

[38] J. Li, J. Chen, R. Ren, X. Cheng, X. Zhao, J.-Y. Nie, and J.-R. Wen. The Dawn After the Dark: An Empirical Study on Factuality Hallucination in Large Language Models. In *Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pp. 10879–10899. Association for Computational Linguistics, Bangkok, Thailand, 2024. doi: 10.18653/v1/2024.acl-long.586 2

[39] S. Li, H. Schieber, N. Corell, B. Egger, J. Kreimeier, and D. Roth. GBOT: Graph-Based 3D Object Tracking for Augmented Reality-Assisted Assembly Guidance. In *2024 IEEE Conference Virtual Reality and 3D User Interfaces (VR)*, pp. 513–523. IEEE, Orlando, FL, USA, Mar. 2024. doi: 10.1109/VR58804.2024.00072 1, 2

[40] Z. Li, H. Zhang, C. Peng, and R. Peiris. Exploring Large Language Model-Driven Agents for Environment-Aware Spatial Interactions and Conversations in Virtual Reality Role-Play Scenarios. In *2025 IEEE Conference Virtual Reality and 3D User Interfaces (VR)*, pp. 1–11. IEEE, Saint Malo, France, Mar. 2025. doi: 10.1109/VR59515.2025.00025 2

[41] E. Marino, L. Barbieri, F. Bruno, and M. Muzzupappa. Assessing user performance in augmented reality assembly guidance for industry 4.0 operators. *Computers in Industry*, 157-158:104085, May 2024. doi: 10.1016/j.compind.2024.104085 1

[42] J. W. Mauchly. Significance Test for Sphericity of a Normal n-Variate Distribution. *Ann. Math. Statist.*, 11(2):204–209, June 1940. doi: 10.1214/aoms/1177731915 6

[43] A. G. Moore, T. D. Do, N. Ruozzi, and R. P. McMahan. Identifying Virtual Reality Users Across Domain-Specific Tasks: A Systematic Investigation of Tracked Features for Assembly. In *2023 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*, pp. 396–404. IEEE, Sydney, Australia, Oct. 2023. doi: 10.1109/ISMAR59233.2023.00054 2

[44] T. Olausson, A. Gu, B. Lipkin, C. Zhang, A. Solar-Lezama, J. Tenenbaum, and R. Levy. LINC: A Neurosymbolic Approach for Logical Reasoning by Combining Language Models with First-Order Logic Provers. In *Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing*, pp. 5153–5176. Association for Computational Linguistics, Singapore, 2023. doi: 10.18653/v1/2023.emnlp-main.313 2

[45] L. Pietschmann, P. Zürcher, E. Bubik, Z. Chen, H. Pfister, and T. Bohné. Quantifying the Impact of XR Visual Guidance on User Performance Using a Large-Scale Virtual Assembly Experiment. In *2023 IEEE Visualization and Visual Analytics (VIS)*, pp. 211–215. IEEE, Melbourne, Australia, Oct. 2023. doi: 10.1109/VIS54172.2023.00051 2

[46] W. Pouw, G. Rop, B. De Koning, and F. Paas. The Cognitive Basis for The Split-attention Effect, Jan. 2019. doi: 10.31234/osf.io/z2q5j 2, 9

[47] U. Radhakrishnan, F. Chinello, and K. Koumaditis. Immersive Virtual Reality Training : Three Cases from the Danish Industry. In *2021 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW)*, pp. 1–5. IEEE, Lisbon, Portugal, Mar. 2021. doi: 10.1109/VRW52623.2021.00008 2

[48] C. Reining, F. Niemann, F. Moya Rueda, G. A. Fink, and M. Ten Hompel. Human Activity Recognition for Production and Logistics—A Systematic Literature Review. *Information*, 10(8):245, July 2019. doi: 10.3390/info10080245 2

[49] P. Sasikumar, R. Hajika, K. Gupta, T. S. Gunasekaran, Y. S. Pai, H. Bai, S. Nanayakkara, and M. Billinghurst. A User Study on Sharing Physiological Cues in VR Assembly Tasks. In *2024 IEEE Conference Virtual Reality and 3D User Interfaces (VR)*, pp. 765–773. IEEE, Orlando, FL, USA, Mar. 2024. doi: 10.1109/VR58804.2024.00096 2

[50] H. Schieber, S. Li, N. Corell, P. Beckerle, J. Kreimeier, and D. Roth. ASDF: Assembly State Detection Utilizing Late Fusion by Integrating 6D Pose Estimation. In *2024 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*, pp. 190–199. IEEE, Bellevue, WA, USA, Oct. 2024. doi: 10.1109/ISMAR62088.2024.00033 2

[51] J.-e. Shin, B. Yoon, D. Kim, and W. Woo. The Effects of Spatial Complexity on Narrative Experience in Space-Adaptive AR Storytelling. *IEEE Trans. Visual. Comput. Graphics*, 29(12):5137–5148, Dec. 2023. doi: 10.1109/TVCG.2022.3201934 2

[52] A. D. Souchet, S. Philippe, D. Lourdeaux, and L. Leroy. Measuring Visual Fatigue and Cognitive Load via Eye Tracking while Learning with Virtual Reality Head-Mounted Displays: A Review. *International Journal of Human–Computer Interaction*, 38(9):801–824, May 2022. doi: 10.1080/10447318.2021.1976509 2, 9

[53] A. Stanescu, P. Mohr, F. Thaler, M. Kozinski, L. R. Skreinig, D. Schmalstieg, and D. Kalkofen. Error Management for Augmented Reality Assembly Instructions. In *2024 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*, pp. 690–699. IEEE, Bellevue, WA, USA, Oct. 2024. doi: 10.1109/ISMAR62088.2024.00084 1, 2

[54] A. Steinfeld, O. C. Jenkins, and B. Scassellati. The oz of wizard: simulating the human for interaction research. In *Proceedings of the 4th ACM/IEEE international conference on Human robot interaction*, pp. 101–108. ACM, La Jolla California USA, Mar. 2009. doi: 10.1145/1514095.1514115 6

[55] M. Toussaint and M. Lopes. Multi-bound tree search for logic-geometric programming in cooperative manipulation domains. In *2017 IEEE International Conference on Robotics and Automation (ICRA)*, pp. 4044–4051. IEEE, Singapore, Singapore, May 2017. doi: 10.1109/ICRA.2017.7989464 1

[56] T. Q. Tran, T. Langlotz, and H. Regenbrecht. A Survey On Measuring Presence in Mixed Reality. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*, pp. 1–38. ACM, Honolulu HI USA, May 2024. doi: 10.1145/3613904.3642383 2

[57] J. Ulmer, S. Braun, C.-T. Cheng, S. Dowey, and J. Wollert. Gamification of virtual reality assembly training: Effects of a combined point and level system on motivation and training results. *International Journal of Human-Computer Studies*, 165:102854, Sept. 2022. doi: 10.1016/j.ijhcs.2022.102854 2

[58] L. Wang, X. Li, J. Wu, D. Zhou, I. Sio Kei, and V. Popescu. AVICol: Adaptive Visual Instruction for Remote Collaboration Using Mixed Reality. *International Journal of Human–Computer Interaction*, 41(2):1260–1279, Jan. 2025. doi: 10.1080/10447318.2024.2313920 1

[59] X. Wang, S. K. Ong, and A. Y. C. Nee. A comprehensive survey of augmented reality assembly research. *Adv. Manuf.*, 4(1):1–22, Mar. 2016. doi: 10.1007/s40436-015-0131-4 1

[60] Z. Wang, F. Kennel-Maushart, Y. Huang, B. Thomaszewski, and S. Coros. A Temporal Coherent Topology Optimization Approach for Assembly Planning of Bespoke Frame Structures. *ACM Trans. Graph.*, 42(4):1–13, Aug. 2023. doi: 10.1145/3592102 2

[61] M. Weib, K. Angerbauer, A. Voit, M. Schwarzl, M. Sedlmair, and S. Mayer. Revisited: Comparison of Empirical Methods to Evaluate Visualizations Supporting Crafting and Assembly Purposes. *IEEE Trans. Visual. Comput. Graphics*, 27(2):1204–1213, Feb. 2021. doi: 10.1109/TVCG.2020.3030400 2

[62] J. W. Woodworth and C. W. Borst. Design and Validation of a Library of Active Affective Tasks for Emotion Elicitation in VR. In *2024 IEEE Conference Virtual Reality and 3D User Interfaces (VR)*, pp. 398–407. IEEE, Orlando, FL, USA, Mar. 2024. doi: 10.1109/VR58804.2024.00061 2

[63] R. Yonetani, T. Taniai, M. Barekatain, M. Nishimura, and A. Kanezaki. Path Planning using Neural A* Search, July 2021. arXiv:2009.07476 [cs]. doi: 10.48550/arXiv.2009.07476 4

[64] R. Zass and A. Shashua. Probabilistic graph and hypergraph matching. In *2008 IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1–8. IEEE, Anchorage, AK, USA, June 2008. doi: 10.1109/CVPR.2008.4587500 2

[65] K. Zhou, Z. Cheng, H. P. H. Shum, F. W. B. Li, and X. Liang. STGAE: Spatial-Temporal Graph Auto-Encoder for Hand Motion Denoising. In *2021 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*, pp. 41–49. IEEE, Bari, Italy, Oct. 2021. doi: 10.1109/ISMAR52148.2021.00018 2

[66] K. Zhou, H. P. H. Shum, F. W. B. Li, and X. Liang. Multi-Task Spatial-Temporal Graph Auto-Encoder for Hand Motion Denoising. *IEEE Trans. Visual. Comput. Graphics*, 30(10):6754–6769, Oct. 2024. doi: 10.1109/TVCG.2023.3337868 2

[67] T. Zigart and S. Schlund. Ready for Industrial Use? A User Study of Spatial Augmented Reality in Industrial Assembly. In *2022 IEEE International Symposium on Mixed and Augmented Reality Adjunct (ISMAR-Adjunct)*, pp. 60–65. IEEE, Singapore, Singapore, Oct. 2022. doi: 10.1109/ISMAR-Adjunct57072.2022.00022 2