The International Journal of Advanced Manufacturing Technology
https://doi.org/10.1007/s00170-026-18707-0

ORIGINAL ARTICLE

A human-centered cognitive support framework for manual industrial assembly: integrating perception, agentic reasoning, and augmented reality guidance

Mariannys Rodriguez[1] · **Efrain Rodriguez**[2] · **Sanderson César Macêdo Barbalho**[1]

Received: 24 May 2026 / Accepted: 7 July 2026

Abstract
Manual industrial assembly remains essential in high-variety and customized production, but increasing product and process complexity places substantial cognitive demands on operators. Existing assistance systems, particularly augmented reality-based solutions, improve instruction visualization and task guidance, yet they often remain weakly connected to the real assembly state and limited in reasoning, adaptation, and decision support. This paper proposes a human-centered cognitive support framework for manual industrial assembly that integrates perception, agentic reasoning, knowledge grounding, and augmented reality guidance. The study is informed by a systematic search, bibliometric overview, and literature analysis of 129 Scopus-indexed documents. The analysis shows that augmented reality dominates current cognitive support approaches, while AI-based methods are increasingly used for object recognition, contextual interpretation, adaptive information delivery, and error detection. However, perception, reasoning, guidance, and knowledge grounding are still commonly treated as isolated functions. Based on these findings, 11 review-derived design requirements are formulated and used to develop a conceptual perception-cognition-guidance framework that represents cognitive support as a closed human-centered loop grounded in procedures, constraints, rules, and memory. The framework is then translated into a layered implementation architecture comprising physical, perception, cognitive, guidance, knowledge, and application layers. Structured data contracts clarify how sensory and interaction data can be transformed into perception evidence, structured assembly states, cognitive support decisions, and device-specific guidance commands. A toy-train assembly demonstrator illustrates how procedural state modeling, multi-camera perception, YOLO11-based object detection, projector-based guidance, and contract-based data exchange can connect physical assembly events with structured reasoning and operator-facing feedback.

Keywords Manual industrial assembly · Human-centered manufacturing · Computer vision · Agentic AI · Augmented reality

1 Introduction

Industrial assembly remains one of the most human-intensive activities in modern manufacturing systems. Despite significant advances in automation and robotics, including collaborative human-robot assembly approaches proposed under the Industry 5.0 paradigm [1], a large portion of assembly operations continue to rely on manual work due to high product variability, frequent design changes, small batch production, and the need for human dexterity and judgment [2]. In many industrial contexts, the deployment of fully automated or robotic assembly solutions is neither economically justified nor technically feasible, particularly

✉ Mariannys Rodriguez
gasca.mariannys@aluno.unb.br

Efrain Rodriguez
efrainrg2009@gmail.com

Sanderson César Macêdo Barbalho
sandersoncesar@unb.br

1 Department of Mechanical and Mechatronics Engineering, University of Brasilia, Campus Darcy Ribeiro, 70910-900 Brasilia, DF, Brazil

2 Department of Mechanical and Mechatronics Engineering, Technological University of Bolivar, Campus Tecnológico Km 1 Vía Turbaco, Cartagena 131001, Bolivar, Colombia

Published online: 13 July 2026

for customized products and complex or low-volume manufacturing. As a result, human operators remain the primary physical actuators in assembly processes.

The transition toward Industry 5.0 further reinforces the central role of humans in manufacturing systems by emphasizing human-centricity, resilience, and sustainable production [3]. Rather than pursuing full automation, Industry 5.0 promotes collaborative and adaptive systems that enhance human capabilities while preserving flexibility and decision-making autonomy [4]. Within this paradigm, the objective is not to replace human labor but to augment human perception, cognition, and performance through intelligent support technologies [5].

Current assistance systems for manual industrial assembly primarily provide procedural guidance, often through static work instructions, digital manuals, or augmented reality overlays [6]. While these approaches can reduce training time and improve task comprehension, they generally lack awareness of the real assembly state, offer limited adaptability to operator behavior or process deviations, and provide minimal cognitive reasoning or decision support. In most cases, guidance remains predefined and reactive, failing to account for unexpected errors, variations in component positioning, or differences in operator expertise

Recent advances in artificial intelligence and sensing technologies have enabled more sophisticated perception and interaction capabilities, including real-time object recognition, pose estimation, and multimodal interfaces [7]. However, existing solutions typically address isolated aspects of the assembly process, such as visual tracking or instructional visualization, without integrating higher-level reasoning, contextual understanding, and adaptive decision-making into a coherent system architecture. Consequently, human operators continue to bear the cognitive burden of diagnosing errors, interpreting complex procedures, and determining corrective actions.

From a human-centered perspective, effective support for manual assembly should extend beyond task visualization toward cognitive augmentation. Such systems should continuously interpret the assembly state, reason over procedural knowledge, adapt guidance to situational and individual factors, and assist decision-making while keeping humans in control of physical execution. Achieving this level of support requires not only advanced perception technologies but also structured mechanisms for knowledge representation, reasoning, memory, and coordinated system behavior.

Agentic generative artificial intelligence offers promising capabilities for high-level reasoning, contextual interpretation, and adaptive interaction, while augmented reality provides a natural interface for delivering context-aware guidance directly within the operator's workspace [8]. Nevertheless, the systematic integration of these technologies into human-centered assembly support systems remains insufficiently developed. Existing approaches still tend to treat perception, reasoning, guidance, and knowledge grounding as separate functions rather than as coordinated elements of a closed cognitive support loop.

This study addresses this gap by proposing a human-centered cognitive support framework for manual industrial assembly. The framework is informed by a systematic functional analysis of 129 documents retrieved from Scopus and is structured to integrate perception, agentic reasoning, and augmented reality guidance. First, the literature is analyzed to identify how existing systems address cognitive support functions, including context awareness, decision support, adaptive guidance, knowledge grounding, and human-centered interaction. Based on this analysis, a set of review-derived design requirements is formulated. These requirements are then used to develop a perception-cognition-guidance framework that conceptualizes cognitive support as a closed human-in-the-loop process grounded in procedures, constraints, and memory. Finally, the framework is translated into a layered implementation architecture, including physical, perception, cognitive, guidance, knowledge, and application layers, supported by structured data contracts[1] and preparation/runtime operation modes.

The main contributions of this paper are as follows:

- A systematic functional analysis of cognitive support in manual industrial assembly, based on 129 documents retrieved from Scopus.
- A set of review-derived design requirements that link limitations in the literature to functional implications for human-centered cognitive support systems.
- A conceptual perception-cognition-guidance framework that represents cognitive support as a closed human-centered loop grounded in procedures, constraints, and memory.
- A layered implementation architecture that operationalizes the conceptual framework through physical, perception, cognitive, guidance, knowledge, and application layers, including structured data contracts and preparation/runtime operation modes.

The remainder of this paper is organized as follows. Section 2 presents the conceptual background on manual industrial assembly, Industry 5.0, and generative and agentic AI in manufacturing. Section 3 describes the systematic search,

[1] In this paper, data contracts refer to explicit structured data objects that define what information is produced by one architectural layer and consumed by another. They are used as interface specifications for traceability and modular integration, not as legal or commercial agreements.

bibliometric overview, and literature analysis. Section 4 derives the design requirements for human-centered cognitive support in manual assembly. Section 5 introduces the proposed conceptual framework. Section 6 presents the layered implementation architecture, including structured data contracts, the toy-train assembly demonstrator, the experimental perception-cognition-guidance sequence, preparation/runtime operation modes, and continuous improvement mechanisms. Section 7 discusses application scenarios, implementation implications, and future research directions. Section 8 concludes the paper.

2 Conceptual background

2.1 Manual industrial assembly in industry 5.0

Modern manufacturing environments are increasingly shaped by product variety, shorter product life cycles, customization demands, and frequent changes in production requirements [9, 10]. These conditions have a direct impact on assembly operations, which remain a critical stage of manufacturing because they transform individual components into functional products through ordered sequences of manual, semi-automated, or automated tasks. Effective assembly planning can reduce process costs, execution time, unnecessary movements, and operational inefficiencies [11].

Assembly also represents a significant share of manufacturing effort. It has been estimated that assembly activities can account for up to 50% of product development time, approximately 20% of total manufacturing time [12], and between 20% and 30% of total production costs [13]. Despite advances in automation, many assembly operations involving high variability, product diversity, and complex manipulation continue to rely heavily on human operators. Current estimates indicate that only 23% of assembly systems are highly automated, while 43% remain predominantly manual and 34% involve hybrid human-machine collaboration [14]. These figures confirm that manual assembly remains a central component of modern production systems.

The persistence of manual assembly is not merely a consequence of limited automation. Human operators provide capabilities that are still difficult to replicate fully through automated systems, including dexterity, flexibility, adaptability, situated decision-making, and problem-solving [15]. These capabilities are particularly valuable in low-volume, high-mix, customized, or frequently changing production environments. However, the same conditions that make human involvement valuable also increase the cognitive demands placed on operators. Manual assembly frequently involves cognitive workload, operator fatigue, training complexity, memory demands, interruptions, and the risk of assembly errors [16]. As product and process variability increase, operators must interpret multiple sources of information, monitor task progress, adapt to unexpected situations, and make decisions under uncertainty.

Recent work on cognition-augmented assembly (CAA) [2] has further clarified that cognitive demands in assembly can be associated with spatiality, memory, knowledge, and decision-making. In practical terms, operators must determine where the relevant object or location is, which parts or resources are involved, how the current assembly state should be interpreted, and what action to perform next. This perspective reinforces the need for cognitive support systems that integrate visual perception, procedural knowledge, and decision support, rather than merely presenting work instructions.

Industry 5.0 provides an appropriate conceptual lens for addressing these challenges because it shifts attention from technology-centered automation toward human-centered, resilient, and sustainable production systems [17]. In this paradigm, the goal is not to remove humans from production, but to design intelligent technologies that augment human capabilities while preserving autonomy, flexibility, and meaningful control. Flexible and reconfigurable assembly systems are therefore increasingly expected to support both operational efficiency and human-centered performance in contexts characterized by mass customization and rapid product changes [18].

Within this perspective, manual assembly should be understood not only as a physical execution task, but also as a cognitive activity. Operators must perceive relevant task conditions, understand procedures, remember sequences, detect deviations, and decide how to act in specific situations. Consequently, support systems for manual assembly should not be limited to displaying instructions. They should help operators build awareness of the assembly state, reason about the task, reduce unnecessary cognitive burden, and receive guidance that is adaptive to the current context. This motivates the need for cognitive support frameworks that integrate perception, reasoning, guidance, and knowledge grounding around the human operator.

2.2 Generative and agentic AI for cognitive support

The evolution of manufacturing technologies has progressively changed both the level of automation and the role assigned to human work. As summarized in Fig. 1, manufacturing systems have moved from mechanization and machine-level automation toward cyber-physical, connected, and increasingly AI-enabled environments. This trajectory has also shifted the role of the operator from

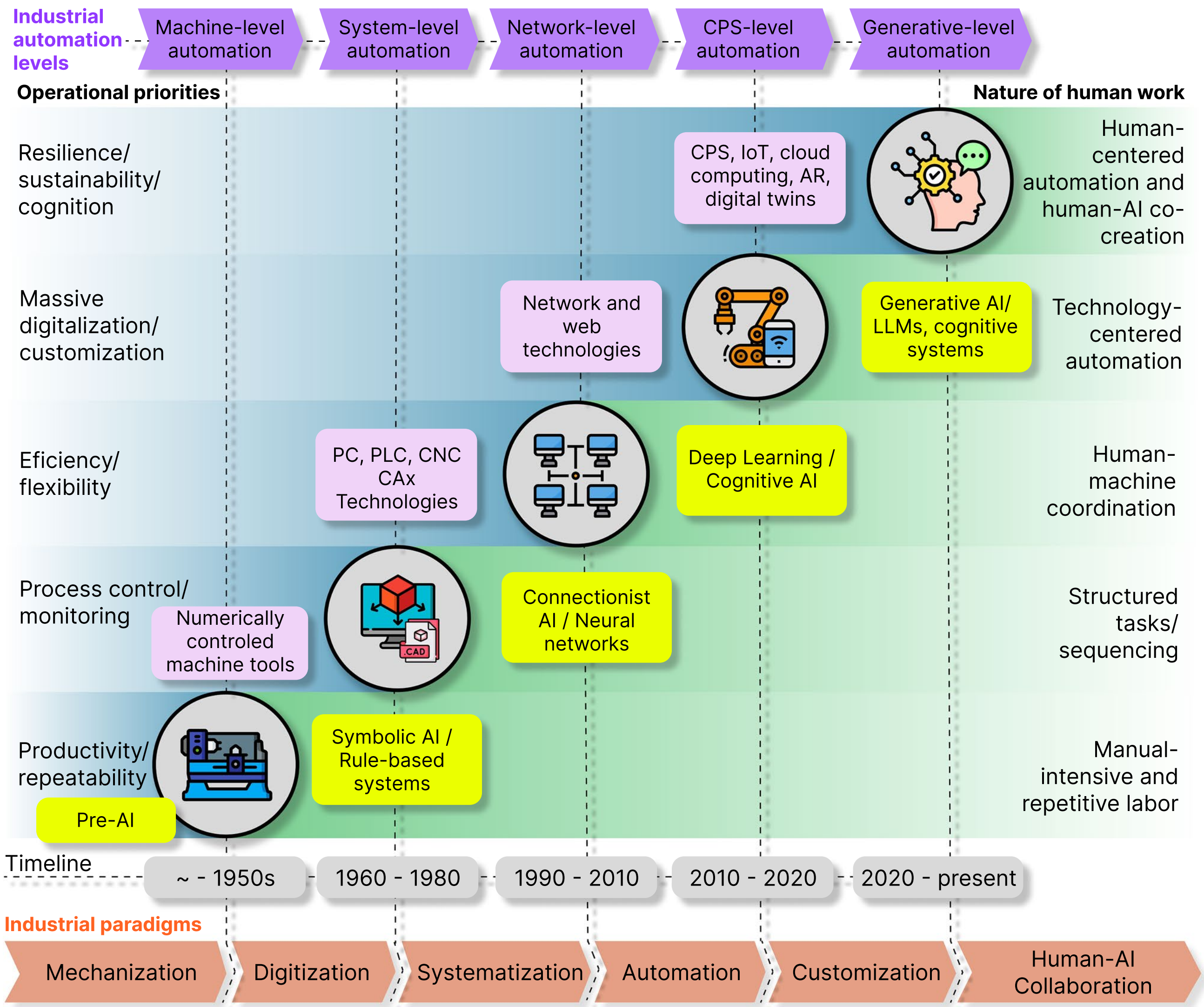


Fig. 1 Evolution of industrial manufacturing systems: from mechanization to human-AI collaboration/co-creation

repetitive execution toward supervision, collaboration, and human-AI co-creation. The figure is not intended to provide a complete historical taxonomy, but to contextualize the broader transition from technology-centered automation to human-centered intelligent support.

Recent advances in artificial intelligence are accelerating this transition. While earlier industrial AI applications were mainly associated with rule-based optimization, predictive analytics, and process automation, current developments in foundation models, generative AI, and multimodal learning are enabling systems with stronger capabilities for contextual interpretation, natural interaction, and adaptive decision support [19, 20]. This evolution is particularly relevant for manual assembly, where support systems must not only present information but also interpret task conditions, respond to operator needs, and adapt guidance to dynamic situations.

Large language models (LLMs) and vision-language models (VLMs) are central to this transition [21]. LLMs provide capabilities for natural language understanding, explanation generation, task reasoning, and interaction with external tools or knowledge bases. VLMs extend these capabilities to visual contexts by linking images or video frames with language-based interpretation, allowing systems to reason about objects, spatial relations, actions, and scene states [22]. In manual assembly, this distinction is important: VLMs can transform visual observations into contextual task representations, while LLMs can support higher-level reasoning over procedures, constraints, operator intent, and decision-making.

Beyond generative content production, these models are increasingly being used as the basis for agentic AI [23]. Agentic AI refers to goal-directed AI systems that combine foundation models with planning, memory, tool use, and iterative reasoning capabilities [8]. In contrast to conventional AI modules that execute predefined tasks, agentic systems can coordinate multiple tools, retrieve relevant knowledge, evaluate intermediate results, and decide on the next action. This makes them relevant for cognitive support in assembly, where the system may need to interpret incomplete visual evidence, compare the current task state with procedural knowledge, determine whether an operator requires assistance, and generate an appropriate response.

It is useful to distinguish agentic AI from classical multi-agent systems (MAS). Traditional MAS have long been used in manufacturing to distribute decision-making among specialized agents for scheduling, monitoring, diagnostics, and control. Their strength lies in modularity, coordination, and distributed autonomy. However, classical MAS are often constrained by predefined rules, fixed communication protocols, or limited contextual reasoning. Recent hybrid approaches argue that LLM-based agentic AI can complement MAS by providing higher-level reasoning, planning, memory, and orchestration, while specialized agents or smaller local models perform domain-specific tasks more efficiently [24]. For manual assembly support, this suggests an architecture in which perception agents, reasoning agents, knowledge agents, and guidance agents can be coordinated within a human-centered support loop.

This perspective aligns with the principles of Industry 5.0, where intelligent systems are expected to augment rather than replace human capabilities. In the context of this work, agentic AI is not treated as an autonomous controller of the assembly process, but as a reasoning and orchestration mechanism that supports the operator. Its role is to connect perceived assembly states, procedural knowledge, operator intent, and guidance generation. Therefore, the relevance of generative and agentic AI for manual assembly lies not merely in generating instructions, but in enabling context-aware, knowledge-grounded, and adaptive cognitive support.

3 Research methodology

The review was designed to provide an evidence base for deriving functional requirements for human-centered cognitive support in manual industrial assembly. Rather than conducting a purely descriptive literature review, the method combines a systematic database search, bibliometric mapping, and qualitative content analysis. The process was informed by the PRISMA statement [25] and is summarized in Fig. 2. The objective was to identify how the literature addresses cognitive support functions such as assembly-state awareness, workload reduction, information presentation, decision support, adaptive guidance, and human-centered interaction.

The search was conducted in Scopus, which was selected for its broad coverage of peer-reviewed research in engineering, manufacturing, and human factors. The following search string was used:

```
TITLE-ABS-KEY(("manual assembly" OR "
industrial assembly"
OR "hand assembly" OR "human assembly")
AND (cognitive OR cognition))
AND NOT TITLE-ABS-KEY(robot* OR "human-
robot")
```

The exclusion of robot-focused terms was applied to maintain the scope of the review on manual and human-centered assembly support, rather than on human-robot collaboration or robotic assembly. The initial query returned 165 records. A publication-year filter from 2015 to 2026 was applied to capture recent developments in cognitive assistance, augmented reality, and AI-enabled manufacturing support. Additional filters restricted the corpus to peer-reviewed articles, review papers, and conference papers. After these filters, 129 documents were retained for analysis.

The retained records were organized in a collaborative spreadsheet including title, authors, year, source, country, keywords, and abstract. Bibliometric analysis was first used to obtain a synthetic overview of the corpus, including publication trends, geographical distribution, and keyword co-occurrence patterns. This step supported the identification of dominant themes and emerging research clusters before the qualitative analysis.

Subsequently, a content analysis was conducted to extract evidence relevant to the research objective. The analysis focused on the problems addressed by the selected studies, the cognitive dimensions considered, the technologies used for support, the type of assembly context, and the reported benefits and limitations. These extraction categories were progressively refined during the analysis and later used to formulate the design requirements presented in Section 4. Therefore, the review method links the literature corpus to the proposed framework by moving from bibliometric mapping, to functional interpretation, to requirement derivation.

The retained corpus was used for bibliometric characterization, while the qualitative content analysis focused on extracting evidence relevant to cognitive support functions. Therefore, the review should be interpreted as a systematic functional analysis rather than as an exhaustive systematic review of all assembly assistance technologies.

3.1 Bibliometric overview

The bibliometric analysis provides an initial characterization of the selected corpus in terms of publication year, country of origin, and keyword co-occurrence. Figure 3a

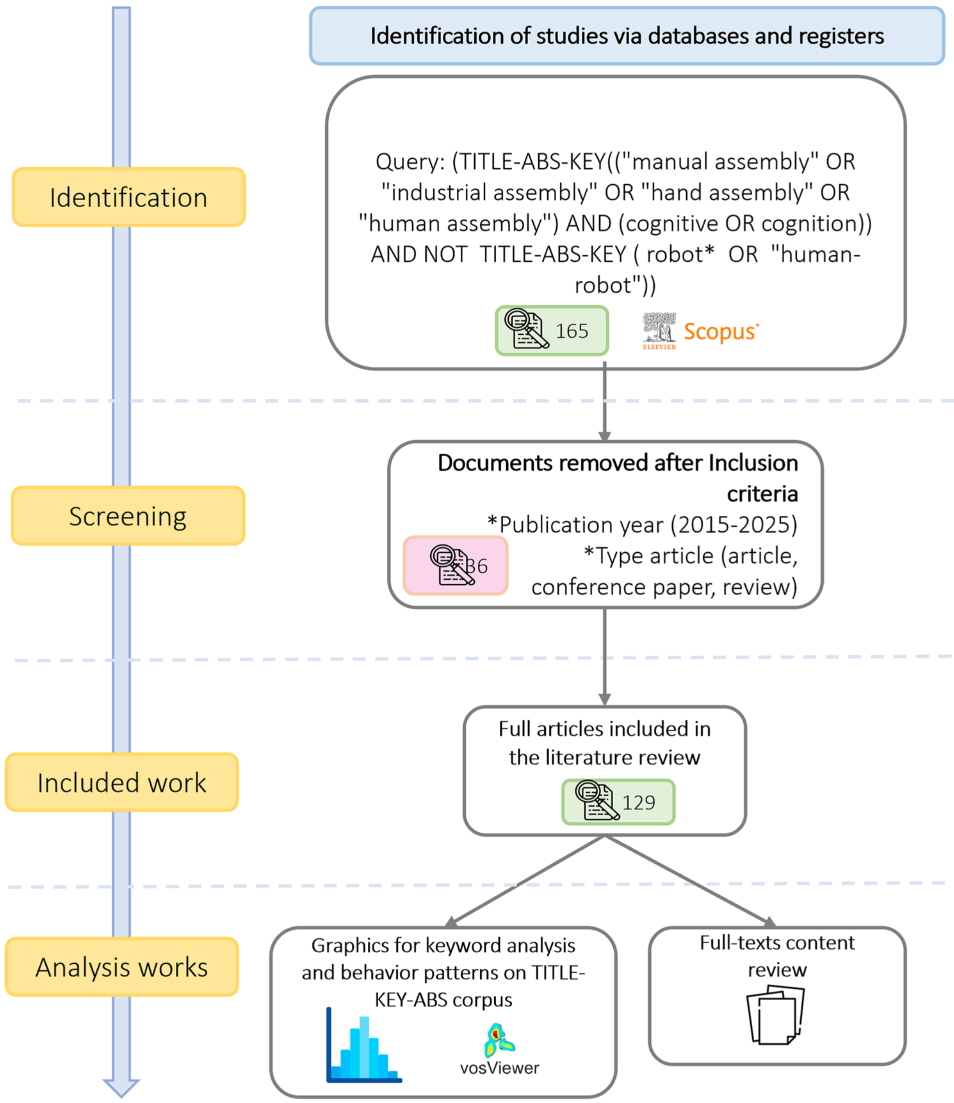


Fig. 2 Overview of the review methodology

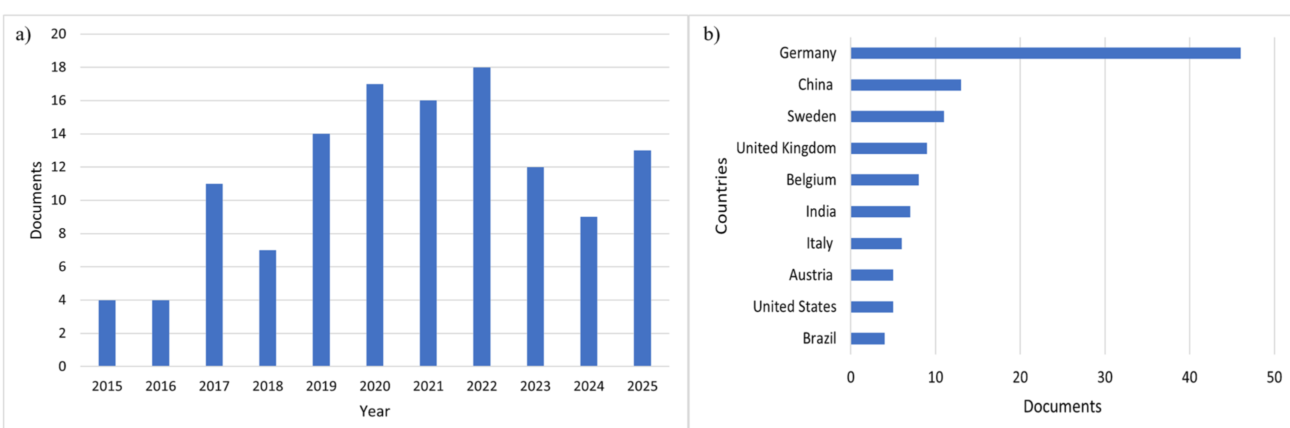


Fig. 3 Statistics from the Scopus database: **a** documents by year; **b** documents by country

illustrates the yearly distribution of publications from 2015 to 2025. During the first years of the analyzed period, publication activity remained relatively limited, suggesting that cognitive aspects of manual assembly were still an emerging research concern. From 2019 onward, the number of publications increased noticeably, with a marked growth between 2020 and 2022. Although a slight decline can be observed in 2023 and 2024, publication output increased again in 2025. Overall, this trend suggests sustained interest in cognitive support for manual assembly, particularly in connection with human-centered manufacturing, digital assistance, and immersive technologies.

Figure 3b shows the geographical distribution of publications by country, considering the ten most productive contributors. The results indicate a concentrated research output, with Germany emerging as the leading contributor. China, Sweden, the United Kingdom, and Belgium also show significant activity, followed by countries such as India, Italy, Austria, the United States, and Brazil. This distribution suggests that research on cognitive support in manual assembly is especially active in countries with strong manufacturing, human-factors, and industrial digitalization research communities.

Figure 4 presents the keyword co-occurrence network obtained from the selected records. A minimum threshold of five keyword occurrences was applied. The network shows assembly as a central node, confirming the relevance of the search strategy to the target domain. Five thematic clusters were identified. The first cluster relates information presentation, cognitive workload, and human performance, indicating the importance of how assembly information is delivered to operators. The second cluster is dominated by augmented reality and its connection with assembly instructions, user interfaces, cognitive load, learning systems, and industrial research. The third cluster focuses on human-centered evaluation, linking task performance, mental workload, human experimentation, and assessment methods such as NASA-TLX. The fourth cluster connects assembly systems, assembly processes, and cognitive ergonomics, emphasizing the relationship between task structure and cognitive demands. Finally, the fifth cluster reflects the emergence of Industry 5.0 and human-centric perspectives in assembly research.

The keyword analysis indicates that immersive technologies, particularly augmented reality, have become central to cognitive support research in manual assembly [26–28].

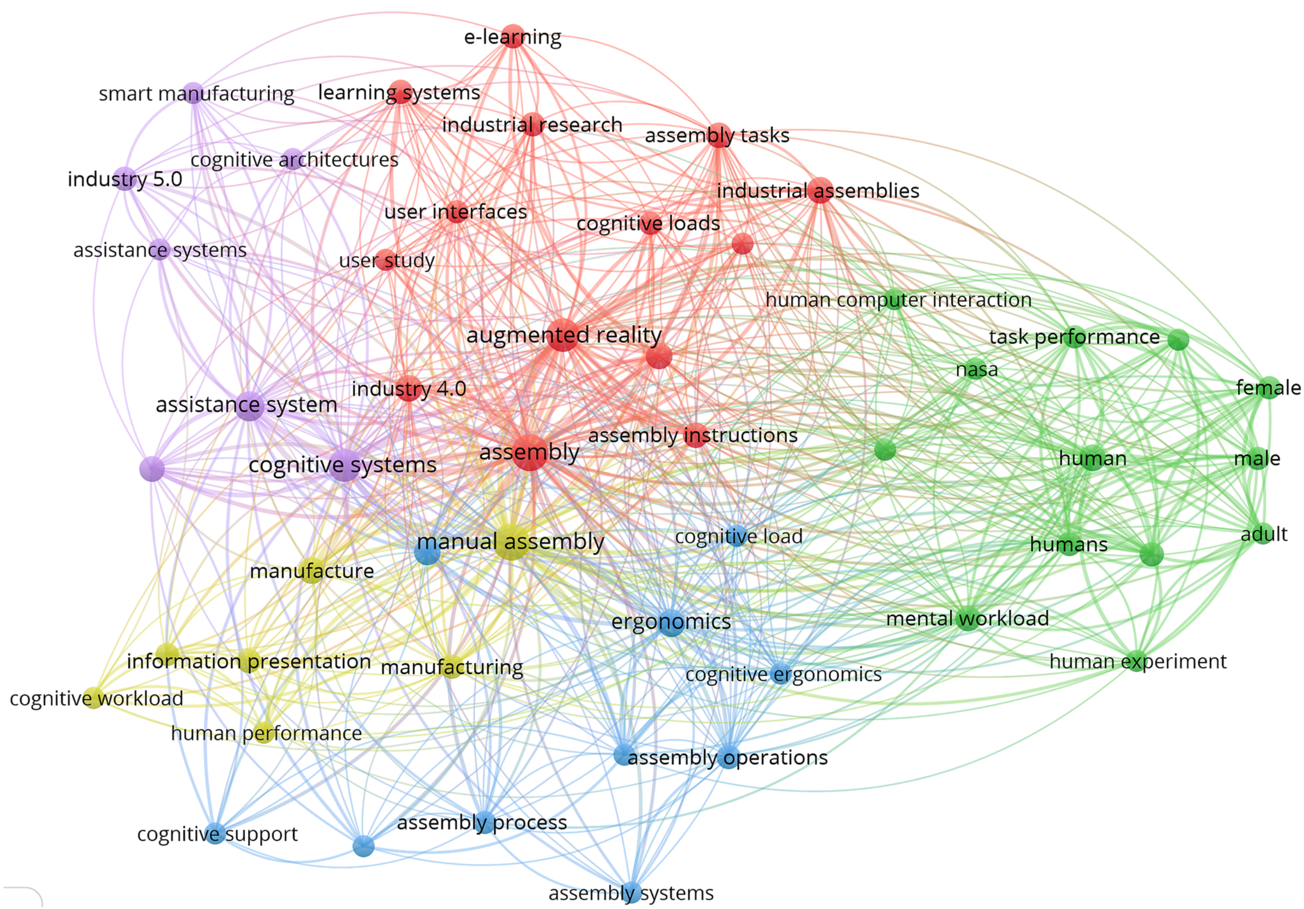


Fig. 4 Keyword co-occurrence network visualization

However, the network also suggests that most contributions remain concentrated on process guidance, visual instruction, user interfaces, and workload evaluation. Cognitive workload and human performance are frequently measured [29–31], but they are less often used as inputs for dynamically adapting the support provided to the operator. Similarly, while Industry 5.0 and human-centered manufacturing appear as emerging themes, the integration of perception, reasoning, adaptive guidance, and knowledge grounding remains limited. This observation supports the need for the functional literature analysis and design requirements developed in the following section.

3.2 Conceptual steps

After the bibliometric overview and qualitative literature analysis presented in Sections 3 and 4, the research evolved toward the conceptual and experimental development of a human-centered cognitive support framework for manual industrial assembly. The methodological progression followed a hypothetical-deductive logic inspired by Karl Popper [32], moving from literature-based observations to conceptual derivation and experimental instantiation.

First, the systematic search, bibliometric mapping, and content analysis were used to identify the main cognitive support dimensions, technological approaches, limitations, and research gaps in manual assembly assistance systems. Based on these findings, Section 4 derived a set of review-based design requirements related to context awareness, adaptive guidance, procedural grounding, human-centered interaction, and cognitive support.

Next, these requirements guided the development of the conceptual perception-cognition-guidance framework presented in Section 5. The framework proposes that cognitive support should operate as a closed human-centered loop integrating perception, reasoning, knowledge grounding, and adaptive guidance.

Subsequently, the conceptual framework was translated into the layered implementation architecture described in Section 6, including perception, cognitive, guidance, knowledge, and application layers, together with structured data contracts connecting the system components.

Finally, the proposed concepts were instantiated in a prototype-oriented experimental setup using the toy-train assembly demonstrator described in Sections 6.2–6.5. The demonstrator illustrates how perception modules, agentic reasoning, and projector-based AR guidance can be connected through structured assembly-state representations and cognitive support decisions. Sections 6.6 and 7 further discuss how the proposed architecture can evolve toward industrial deployment scenarios such as training, production, and maintenance support.

4 Literature analysis and design requirements

This section synthesizes the main findings obtained from the content analysis of the selected literature corpus. The analysis was organized around three complementary dimensions. First, the reviewed studies were examined in terms of the problems they address in manual assembly, with particular attention to assembly complexity and its impact on operator performance. Second, the cognitive support dimensions discussed in the literature were analyzed, including cognitive load, information presentation, attention guidance, learning, decision-making, memory, and error prevention. Third, the technological approaches used to support these dimensions were mapped, with emphasis on XR, AI, digital systems, and sensor-based solutions. This synthesis provides the basis for deriving the design requirements that guide the proposed framework.

4.1 Manual assembly complexity

A recurring theme in the reviewed literature is that contemporary assembly processes are increasingly complex. This complexity is mainly associated with high product variety, short product life cycles, frequent product changes, and small batch sizes [33–35]. Under these conditions, manual assembly often remains the most feasible solution because human operators provide flexibility and adaptability that are difficult to achieve with fully automated systems [36–39]. However, this flexibility has a cost: part of the complexity that cannot be absorbed by automation is transferred to the operator.

As a result, operators are required to manage variability, interpret multiple information sources, coordinate interdependent actions, execute non-linear task sequences, meet quality requirements, and adapt to process changes [40–43]. This increases the demand for operator skills and cognitive capabilities in assembly environments [44–46]. For instance, in high-variety assembly lines, operators may need to memorize new instructions whenever product variants change [47]. Complex products may also involve long sequences, difficult-to-remember requirements, and high quality standards [48]. In addition, traditional instruction media, such as paper-based manuals, separate information from the physical object, which can distract operators and hinder task execution.

Dynamic workplace conditions further intensify these challenges. Interruptions can increase the time required to resume assembly tasks, especially when they are similar to the primary activity [49]. Forgetting effects also deteriorate performance in production environments where product changes and task interruptions are frequent [50].

These factors affect not only individual performance but also operational variables such as assembly time, rework, and quality. Previous studies have reported that higher product complexity is associated with longer assembly times, increased rework rates, and more human errors [33, 46, 51]. In this context, assembly errors remain a critical concern, with a substantial share of quality defects being directly or indirectly associated with human error [28].

Therefore, manual assembly complexity should be understood as a multidimensional phenomenon emerging from the interaction between product, process environment, and operator characteristics [51–54]. This implies that effective support cannot be limited to physical assistance or instruction visualization. It must also address cognitive aspects such as perception, attention, learning, memory, decision-making, and process monitoring [36, 37]. This observation directly motivates the need for support systems capable of representing the assembly state, interpreting task progress, and adapting assistance to the operator and context.

4.2 Dimensions of cognitive support in manual assembly

Given the cognitive demands imposed by assembly complexity, the literature has increasingly examined how assistance systems can support operator cognition. The most recurrent dimensions include cognitive load management [55], information presentation [56, 57], attention guidance, learning and skill development, decision-making support, memory support, and error prevention. Additional factors such as fatigue [58], knowledge management, usability, and operator acceptance [59] also influence the effectiveness of cognitive assistance systems in industrial environments.

Cognitive load management is one of the most frequently discussed dimensions [60]. Excessive information [61], poorly timed instructions [62], interruptions [49, 63, 64] and memory demands [65, 66] can impair assembly performance. Wang et al. [44] reported that both information overload and information underutilization may negatively affect task execution. Similarly, mental workload and frustration have been identified as relevant predictors of fatigue in complex manual assembly tasks [67–69], while cognitive interference in working memory can accelerate forgetting after interruptions [50]. EEG-based studies further suggest that in situ projected instructions can reduce working memory load compared with traditional paper-based instructions [16]. These findings indicate that cognitive support systems should not only provide more information, but also regulate the amount, timing, and format of information delivered to the operator [70].

Information presentation and visualization are closely related to cognitive load [39, 71]. The design of instructional content strongly affects how operators perceive and process assembly information [66, 72–74]. Different presentation methods, including text, photographs, animations, 3D assistance, and projection-based instructions, have different effects on task performance and cognitive economy [41, 75–79]. Workstation design and material presentation strategies, such as numbered boxes or structured layouts, can also reduce stress and workload while improving productivity and accuracy [80, 81]. Product architecture and component design can further shape the cognitive effort required during assembly [82]. These findings suggest that support must be both task-relevant and spatially situated.

Attention guidance and perception represent another important dimension. Projection systems and AR-based interfaces can guide the operator's attention toward relevant parts, tools, or locations, reducing the need to search for information and minimizing distractions [83]. However, the literature also indicates that AR systems still face challenges related to ergonomic maturity, usability, and real-world industrial deployment [83]. This suggests that attention guidance must be designed carefully to avoid increasing cognitive load or reducing awareness of surrounding events.

Learning and skill development are also central, especially in training, onboarding, and high-variety assembly contexts [45, 66, 84, 85]. Skilled and novice operators differ in task speed, competence, and subjective experience [62, 86]. Cognitive assistance systems can support learning by providing step-by-step guidance, reducing dependence on prior experience, and adapting information as operators gain competence. Projection-based and digital assistance systems have shown potential to improve initial assembly cycles and support autonomous learning [87, 88]. Such findings are particularly relevant for training, ramp-up, and new product introduction scenarios.

Decision-making support becomes critical in dynamic or variant-rich environments. Cognitive assistance systems can provide context-specific information to support operator decisions during ongoing assembly processe [89, 90]. This is especially relevant in rework and repair, where process flows may depend on product variants, detected faults, or incomplete information [91]. However, many existing systems remain rigid, requiring strict adherence to predefined procedures and limiting operator flexibility in dynamic contexts [42, 91]. This limitation highlights the need for systems that preserve human autonomy while providing reasoning support.

Memory support and knowledge management are also important because assembly performance is affected by learning and forgetting effects [50, 92]. Assistance systems can act as external cognitive aids by reducing the need to memorize complex procedures and by preserving process knowledge across operators and production changes.

Knowledge management is therefore not merely an organizational concern but a functional requirement for cognitive support in manual assembly [34].

Finally, error prevention and detection are among the most tangible outcomes of cognitive assistance systems [38, 93]. Case studies have reported substantial reductions in assembly errors compared with non-assisted assembly, especially for omission errors [94, 95]. Camera-based inspection systems can also support process verification and traceability [91]. Since human errors are often associated with excessive information search, decision-making burden, and cognitive overload [96], reducing cognitive burden can directly contribute to quality improvement.

Despite these benefits, the literature consistently reports limitations related to usability, worker acceptance, implementation complexity, and industrial validation. Existing systems often lack the simplicity, ergonomic maturity, and robustness required for widespread adoption [83, 97]. Some spatial projection systems may be useful for training and new product introduction, but are not yet perceived as sufficiently mature for serial production [86, 98]. Operator acceptance remains a critical factor, and early worker involvement in system design can improve adoption and perceived usefulness [99]. These limitations reinforce the need for support frameworks that are not only technically capable, but also human-centered, adaptive, and feasible for real industrial settings.

4.3 Technological approaches for cognitive support

The reviewed literature shows that cognitive support in manual assembly has been addressed through a variety of technological approaches, including Extended Reality (XR), Artificial Intelligence (AI), digital systems, and sensor-based technologies. These technologies support various functions, including perception, learning, decision-making, guidance, error prevention, and quality management. However, they are often studied in isolation, which limits the development of integrated support systems. Table 1 summarizes the dominant contribution of each technological family to human-centered cognitive support in manual assembly and highlights the limitations that motivate the proposed framework.

According to a visual-computation perspective [2], these technologies can also be interpreted according to their role in the assembly cognition process. Position registration supports spatial cognition by locating assemblies, parts, tools, or target regions in the workspace. Multi-layer recognition supports memory-related cognition by identifying features, parts, resources, and assembly states at different levels of granularity. Contextual perception supports knowledge-related cognition by interpreting process progress, quality conditions, and potential errors. Mixed-reality fusion supports decision-making by translating computed assembly information into situated guidance. This functional interpretation reinforces the need for an architecture that does not treat computer vision, reasoning, and AR guidance as independent modules, but as connected stages of cognitive support.

Table 1 Simplified mapping of technological approaches for cognitive support in manual assembly

Technological family	Main assembly applications	Primary cognitive support role	Observed limitation / gap
Extended Reality (AR/MR/VR)	Training, assembly guidance, process planning, ergonomic evaluation, remote instruction, inspection	Spatial visualization, attention guidance, situated instructions, interactive feedback	Often focused on predefined visual guidance; limited reasoning, adaptation, and integration with task-state awareness
Artificial Intelligence and Computer Vision	Object recognition, action recognition, error detection, adaptive instructions, decision support	Perception, context interpretation, personalization, anomaly/error awareness, decision support	Frequently implemented as isolated perception or prediction modules, with limited connection to procedural reasoning and guidance delivery
Digital and sensor-based systems	IoT monitoring, digital twins, workstation sensing, traceability, real-time feedback	Data capture, process monitoring, state tracking, operational traceability	Often emphasize data acquisition or monitoring rather than human-centered cognitive support and adaptive interaction
Instructional and interface technologies	Digital work instructions, video assistance, screen-based interfaces, gamified learning	Information presentation, learning support, memory aid, operator engagement	May reduce training effort but often remain static, non-context-aware, and weakly connected to real-time assembly state

XR technologies are the most prominent technological family in studies related to manual assembly and cognition. Within this group, AR is the dominant approach for delivering cognitive support because it can provide visual instructions, spatial cues, and real-time interactive guidance directly in the operator's work environment [100–102]. VR is more frequently used in pre-assembly contexts, such as process planning, operator training [103], simulation, design evaluation, and ergonomic assessment [17, 104–106]. AR and MR, in contrast, are more commonly applied in in-situ and post-assembly scenarios, including assembly guidance

[53, 107–109], task allocation [17], quality inspection [52, 110, 111], error detection [97] and remote instruction[111].

Several studies report benefits of AR-based systems over traditional instruction methods, including improved skill acquisition, user engagement, process flexibility, and reductions in physical or cognitive workload [86, 112, 113]. The integration of VR and AR technologies has also shown potential to improve performance and reduce errors in industrial assembly environments [114]. Nevertheless, AR effectiveness is not universal. Some studies report increased task completion times or cognitive workload in simple assembly scenarios, while AR may offer clearer benefits in more complex tasks [115]. AR systems may also reduce operators' awareness of external events, creating potential safety concerns in real workspaces.

Recent studies increasingly point toward the convergence of AR with AI. AI methods can support context interpretation, adaptive information delivery, personalization, decision-making, and error detection [15, 48, 73, 86, 95, 116]. Deep learning approaches, particularly object recognition and activity recognition, have demonstrated potential for reducing assembly errors and enabling modular cognitive assistance systems [94, 97]. More recent developments in generative and agentic AI extend these possibilities by enabling reasoning, explanation, and orchestration capabilities that can connect perception outputs with procedural knowledge and guidance decisions.

Overall, the analysis reveals a technological division that motivates the proposed framework. XR technologies are strong in situated visualization and interaction, while AI technologies contribute to perception, adaptation, reasoning, and decision support. Digital and sensor-based systems provide monitoring, traceability, and state information, but are not sufficient by themselves to support cognitive interaction. Existing solutions therefore tend to address only partial aspects of the problem. This fragmentation highlights the need for an integrated human-centered framework that connects perception, cognition, knowledge grounding, and guidance into a closed cognitive support loop for manual assembly.

4.4 Review-derived design requirements

The literature analysis indicates that cognitive support in manual assembly cannot be reduced to the visualization of work instructions. The reviewed studies show that assembly complexity is transferred to operators through product variety [51], process variability [117], interruptions, memory demands, quality requirements [118], and the need for situated decision-making [10, 43, 49, 119]. At the same time, existing assistance systems often address isolated functions, such as instruction visualization [120], workload measurement [35, 53], or object recognition, without integrating perception, reasoning, adaptation, and guidance into a unified human-centered support logic. Based on these findings, a set of review-derived design requirements was formulated to guide the development of the proposed framework. These requirements are not intended as exhaustive industrial specifications, but as functional principles that translate the main limitations and needs identified in the literature into framework-level design implications.

The requirements listed in Table 2 were classified according to two complementary criteria: the type of evidence supporting their derivation and their role within the proposed framework. The evidence type distinguishes between recurrent themes repeatedly observed in the reviewed literature, gap-driven needs that emerge from limitations or underdeveloped capabilities in existing systems, paradigmatic requirements associated with the human-centered orientation of Industry 5.0, and implementation-driven concerns related to feasibility, usability, validation, and industrial adoption [121–124]. The framework role indicates whether each requirement is central to the cognitive support logic, mainly related to human-system interaction, or enabling for practical implementation and continuous improvement.

Taken together, these requirements suggest that an effective cognitive support system for manual assembly should operate as a closed human-centered support loop rather than as a one-way instruction delivery mechanism. The system must observe the evolving assembly situation, interpret it against procedural knowledge, decide whether support is necessary, and deliver spatially meaningful guidance while preserving operator autonomy. Therefore, the proposed framework is organized around three functional domains: perception, cognition, and guidance; grounded by assembly knowledge and centered on the human operator.

5 Conceptual framework for human-centered cognitive support

The design requirements derived in the previous section indicate that cognitive support in manual assembly should not be conceived as a one-way delivery of work instructions. Instead, it should operate as a closed human-centered support loop capable of observing the assembly situation, interpreting it against procedural knowledge, deciding whether support is needed, and delivering guidance while preserving operator autonomy. Figure 5 presents the conceptual vision of the proposed framework, organized around three functional domains: perception, cognition, and guidance. These domains represent the minimum set of capabilities required to move from static instruction delivery toward adaptive, context-aware, and cognitively informed support.

Table 2 Design requirements for human-centered cognitive assembly support

ID	Design requirement	Review-derived need	Evidence type	Framework role	Framework implication
DR1	Context-aware assembly state representation	Operators must manage variability, spatial relations, and task progress.	Recurrent	Core	Perception
DR2	Cognitive load-sensitive information delivery	Information overload, interruptions, and fatigue affect performance.	Recurrent	Core	Guidance
DR3	Spatially situated and task-relevant guidance	Static or separated instructions can distract from the physical task.	Recurrent	Interaction	Guidance
DR4	Procedural knowledge grounding	Support must follow valid rules, task sequences, and constraints.	Gap-driven	Core	Knowledge grounding and cognition
DR5	Deviation and error awareness	Errors, omissions, wrong sequences, and quality defects remain critical.	Recurrent	Core	Perception and cognition
DR6	Adaptive support to operator expertise and task evolution	Operators differ in expertise, pace, learning needs, and preferences.	Emerging / gap-driven	Core	Guidance and memory
DR7	Human control and decision autonomy	Human-centered manufacturing requires augmentation, not replacement.	Paradigmatic	Core	Human-centered loop
DR8	Operational traceability and learning from use	Validation, feedback, and continuous improvement require usage evidence.	Implementation-driven	Enabling	Memory and application layer
DR9	Modular and feasible implementation	Industrial adoption depends on usability, ergonomics, and integration.	Implementation-driven	Enabling	Implementation architecture
DR10	Multimodal interaction and input handling	Operators interact through action, speech, text, gestures, and vision.	Emerging / gap-driven	Interaction	Perception, guidance, and human-centered loop
DR11	Traceable reasoning and support justification	Adaptive recommendations require transparency, validation, and trust.	Implementation-driven	Enabling	Cognition, explanation, and logs

At the center of the framework is the assembly human operator, who remains the primary actor in the manual assembly process. This central position reflects a key assumption of the proposed approach: the objective is not to replace human execution or decision-making, but to augment the operator's ability to perceive relevant task conditions, interpret the current situation, and act appropriately. The operator physically performs assembly operations, manipulates parts and tools, responds to system cues, expresses intent or uncertainty, and generates feedback that allows the system to update its understanding of the task.

The three surrounding domains define the core support cycle. The perception domain is responsible for observing and interpreting the assembly state. Supported by computer vision and vision-language models, it captures multimodal information from the workspace, including operator activity, part-tool interactions, object states, spatial relationships, and task progress. Its role is not limited to raw sensing; rather, perception transforms observable activity into contextual information usable by higher-level reasoning processes.

The cognition domain represents the reasoning core of the framework. It is associated with agentic reasoning because this domain requires interpreting perceived context, inferring operator intent, evaluating task conditions, and determining the type of support needed. In this framework, agentic AI is not understood as an unconstrained generator of instructions, but as a reasoning and orchestration mechanism grounded in procedures, constraints, rules, and memory. Its purpose is to connect what is observed in the workspace with what should happen according to the assembly process.

The guidance domain translates cognitive decisions into operator-facing support. Enabled by augmented reality, this domain delivers cues, instructions, warnings, confirmations, or explanations directly within the operator's work context. Guidance is therefore not treated as static visualization, but as an adaptive interaction mechanism that should provide the right support at the right moment, in a form that is spatially meaningful and minimally disruptive to the assembly task.

The circular arrows in Fig. 5 describe the main flow of cognitive support. Perception transforms operator activity and task state into context; cognition converts this context into support through reasoning and decision-making; and guidance produces feedback by influencing the operator's next action and creating a new observable state in the workspace. This loop makes the system adaptive because each operator's response and task outcome becomes part of the next perception-cognition-guidance cycle.

The bidirectional arrows between the operator and the three domains emphasize that the human is not a passive recipient of system output. The connection with perception captures the relation between activity and state, since the operator's actions and the evolving assembly state are

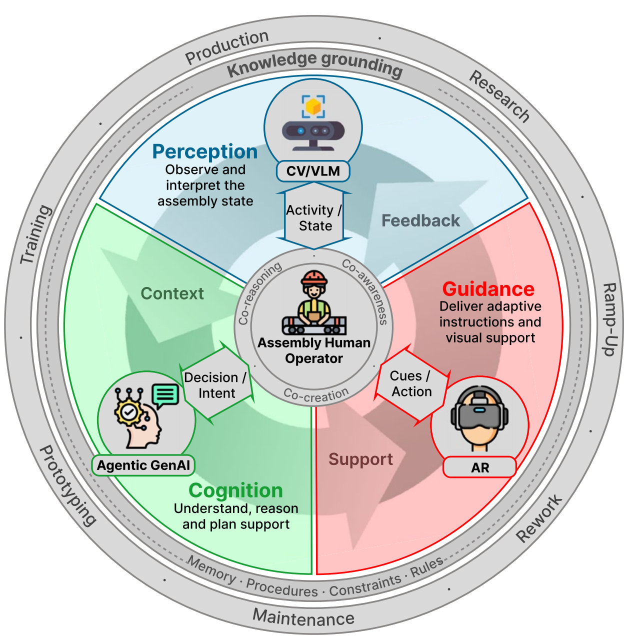


Fig. 5 Human-centered cognitive support framework for manual industrial assembly

continuously observed and interpreted. The connection with cognition reflects decision-making and intent because the system must account for human goals, uncertainty, confirmation, and situated decisions. The connection with guidance reflects cues and actions, in which support is delivered to the operator and translated into physical assembly behavior.

Knowledge grounding indicates that the complete loop must be informed by procedures, constraints, rules, and memory. These elements support perception by defining what is relevant to observe, support cognition by constraining interpretation and reasoning, and support guidance by ensuring that assistance remains valid, safe, and task-relevant. The framework also enables three human-centered interaction principles: co-awareness, through a shared interpretation of the evolving assembly state; co-reasoning, through the interaction between operator intent and system-supported decision-making; and co-creation, through the continuous shaping of task execution by human action, adaptive guidance, and feedback. In this sense, the operator remains in control while the system interprets, reasons, and supports the assembly process.

Finally, the framework is intentionally general enough to be instantiated across different assembly contexts. These include training, production, ramp-up, research, prototyping, and quality-oriented support. These contexts are discussed later as potential use cases, while the following section translates the conceptual framework into a layered implementation architecture.

6 Layered implementation architecture

The conceptual framework introduced in Section 5 defines the functional logic of the proposed cognitive support system: perception, cognition, guidance, knowledge grounding, and human-centered interaction. The review-derived design requirements further indicate that such a system must be context-aware, cognitively adaptive, procedurally grounded, traceable, multimodal, and feasible for industrial implementation. This section translates these principles into a layered implementation architecture that specifies the main hardware elements, software functions, data flows, agentic reasoning components, and operational modes required to instantiate the framework.

Figure 6 presents the proposed architecture. It is organized around six complementary layers: the physical layer, perception layer, cognitive layer, guidance layer, knowledge layer, and application layer. The architecture should be understood as an implementation-oriented reference architecture rather than a fixed software stack. It defines how different components can interact while

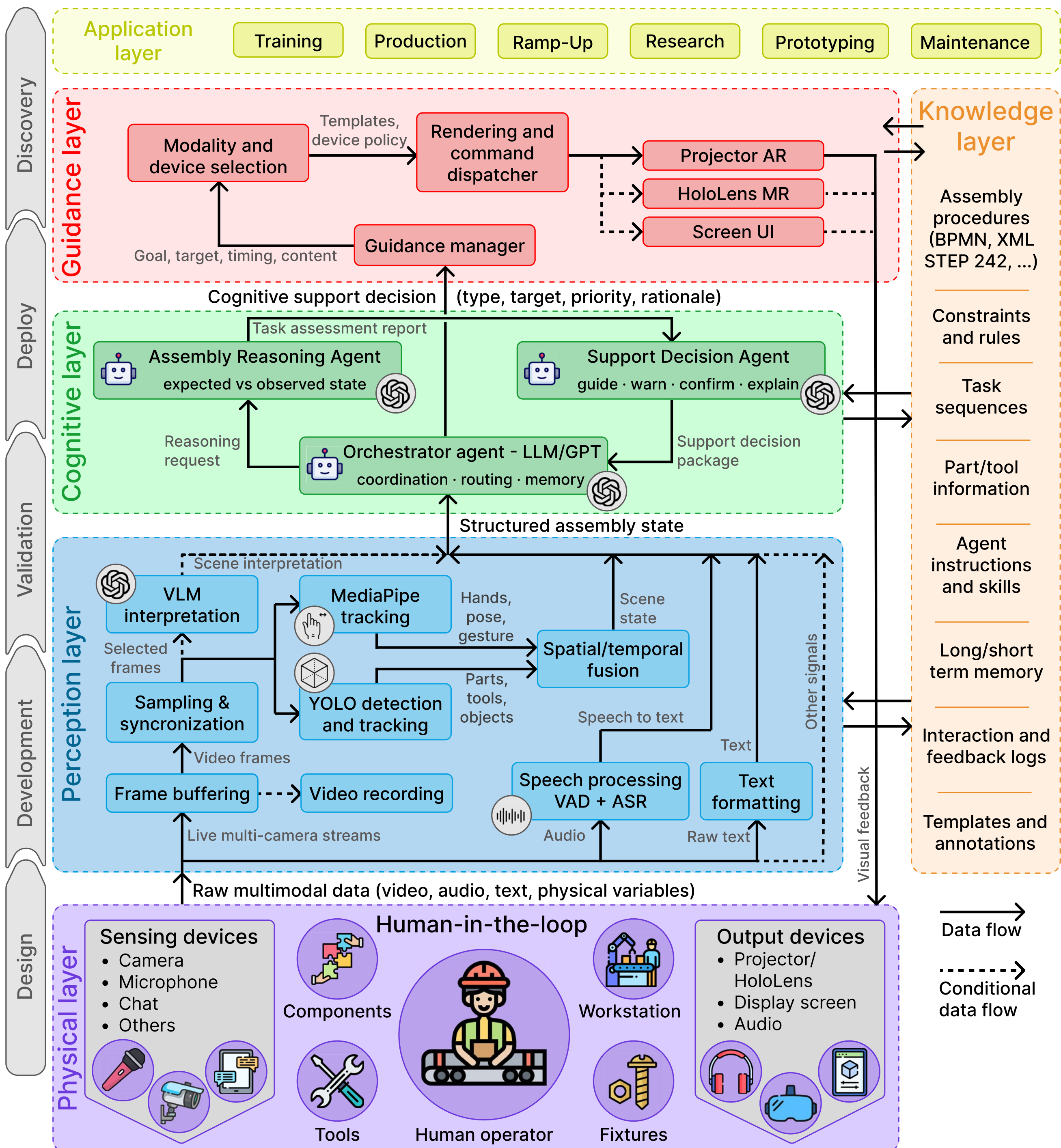


Fig. 6 Layered implementation architecture of the proposed human-centered cognitive support system

allowing alternative models, devices, or platforms to be substituted according to the assembly context.

The physical layer represents the human-centered assembly environment. It includes the human operator, assembly components, tools, fixtures, workstation, and the physical devices required for sensing and feedback. On the input side, cameras, microphones, chat or text interfaces, and additional sensors capture information about the assembly process. On the output side, projector-based AR, HoloLens MR, screen displays, and audio devices deliver situated support to the operator. This layer is not passive: the operator remains the primary physical actor, while the system observes, interprets, and supports the task without removing human decision autonomy.

The perception layer transforms raw multimodal data into a structured representation of the assembly situation. In the visual branch, real-time multi-camera streams are buffered, optionally recorded, sampled, and synchronized before being processed by computer vision modules. YOLO-based detection and tracking can be used to identify parts, tools, workpieces, and relevant assembly objects, while MediaPipe-based tracking can provide hand, pose, and gesture information. These outputs are combined through spatial and temporal fusion to infer object identities, hand-object relations, object-object relations, motion, visibility, and uncertainty. A vision-language model can be invoked using selected frames when semantic scene interpretation is required, particularly under occlusion, ambiguity, or operator questions that depend on the current visual context.

The perception layer also includes audio, text, and additional signal branches. Audio input can be processed through voice activity detection and automatic speech recognition, producing a timestamped transcript rather than a final interpretation of operator intent. Text input from a chat interface, tablet, computer, or AR menu can be formatted and timestamped in a similar way. Other physical variables, such as lighting, temperature, noise, or equipment signals, can be included as additional context. The output of this layer is not a final decision, but a structured assembly state that integrates visual evidence, operator input, environmental signals, confidence values, uncertainty, and references to the evidence used to construct the state.

The cognitive layer receives the structured assembly state and converts it into a cognitive support decision. It is organized as a compact agentic reasoning structure with three main roles. The Orchestrator Agent coordinates the reasoning process, builds the cognitive context package, retrieves relevant knowledge, routes information to the appropriate reasoning functions, and manages short-term memory. The Assembly Reasoning Agent compares the observed assembly state with the expected process model, evaluating task progress, preconditions, deviations, uncertainty, and operator requests. The Support Decision Agent determines whether support is required and selects the support type, target, priority, and rationale. This agentic organization is used only where flexible reasoning is required: perception remains responsible for structuring evidence, while guidance remains responsible for delivering feedback.

The output of the cognitive layer is a decision on cognitive support. This decision specifies what kind of support is needed, why it is needed, where it should be directed, and what priority it should have. It does not prescribe the final rendering format. For example, the cognitive layer may determine that the operator should verify the orientation of a wheel before insertion, but the guidance layer decides whether this support should be shown as a projected highlight, a HoloLens panel, a screen update, an audio cue, or a combination of modalities.

The guidance layer translates the cognitive support decision into operator-facing feedback. It includes a guidance manager, a modality and device selection module, and a rendering and command dispatcher. The guidance manager interprets the support decision in terms of goal, target, timing, and content. The modality and device selection module determines whether the support should be delivered through projector-based AR, HoloLens MR, screen UI, or audio. The rendering and command dispatcher generates the final device-specific commands. This separation decouples reasoning from rendering and allows the same cognitive support decision to be expressed through different guidance technologies depending on the task context, available devices, and operator needs.

The knowledge layer grounds the architecture across all functional layers. It stores assembly procedures, task sequences, constraints, rules, part and tool information, agent instructions and skills, long-term and short-term memory, interaction logs, feedback logs, templates, and annotations. It supports the perception layer by providing task-relevant objects, labels, expected states, and calibration-related context. It supports the cognitive layer by providing procedural knowledge, constraints, and memory. It supports the guidance layer by providing templates, device policies, and annotated targets. In this sense, the knowledge layer is not merely a database, but a grounding mechanism that constrains perception, reasoning, and guidance.

Finally, the application layer defines the deployment context of the architecture, including training, production, ramp-up, research, prototyping, quality support, and maintenance. These contexts influence the level of guidance detail, the interaction policy, the logging strategy, and the acceptable degree of system intervention. The lifecycle arrows on the left of the Fig. 6 frame the architecture as an evolving implementation process rather than a static runtime pipeline. They indicate that the same layered system must be designed, developed, validated, deployed, and maintained, with runtime evidence feeding subsequent refinement of models, agents, knowledge resources, guidance assets, and physical calibration.

6.1 Structured data contracts between layers

To make the implementation architecture operational, each layer exchanges structured data objects rather than relying only on unstructured signals or implicit module outputs. These objects can be understood as representative data contracts between the system's main components. They are not intended to impose a fixed software schema, but rather to clarify the minimum information required for each layer to communicate with the next. In this way, the architecture becomes modular, traceable, and extensible: perception modules can be modified without changing the cognitive agents, cognitive reasoning can be updated without redesigning the rendering devices, and guidance technologies can be replaced without redefining the entire system.

The data-contract logic follows a progressive transformation chain: raw multimodal data are converted into perception evidence; perception evidence is fused into a structured assembly state; the structured assembly state is converted into a cognitive support decision; and the support decision is translated into a device-specific guidance command. This chain preserves traceability from sensing to feedback and supports the design requirement for transparent and auditable reasoning.

Table 3 summarizes the main data objects exchanged across the architecture. These contracts follow the transformation of information from sensing to feedback. Raw multimodal streams are first converted into synchronized visual events, object detections, hand-tracking evidence, operator transcripts, and perception-level evidence. These elements are then consolidated into a structured assembly state that integrates scene relations, operator input, uncertainty, and references to the evidence used to construct the state. The cognitive and guidance layers subsequently use this state to generate reasoning requests, task assessments, support decisions, and device-specific guidance commands. Therefore, the table should be read as an interface map: it identifies what each layer produces, what the next layer consumes, and how traceability is preserved from sensing to operator-facing feedback.

6.2 Toy-train assembly demonstrator

To instantiate the data-contract logic in a concrete assembly scenario, a toy-train assembly task was used as a prototype-oriented demonstrator. The objective of this demonstrator is not to validate the complete cognitive support system under industrial conditions, but to illustrate how a simplified manual assembly process can be represented procedurally, observed through perception modules, and connected to projector-based guidance through structured data objects. The demonstrator therefore provides a controlled case for showing how the proposed architecture can move from physical assembly activity to perception evidence, structured assembly states, and support-oriented feedback.

Table 3 Representative structured data contracts exchanged across the implementation architecture

Data Object	Produced by	Consumed by	Purpose
`rawmultimodaldata`	Physical layer	Perception layer	Registers video, audio, text, and sensor streams.
`visionprocessingevent`	Camera acquisition	Visual perception	Describes sampled and synchronized camera frames.
`objectdetectiontracking`	YOLO11 + tracking	Fusion module	Stores detected parts, tools, identity, confidence, and motion.
`handposetracking`	MediaPipe	Fusion module	Stores hand landmarks, gestures, and hand-object proximity.
`perceptionevidence`	Visual branch	Fusion module	Aggregates detections, tracks, hand evidence, and view metadata.
`operatortranscript`	Speech processing	Cognitive layer	Provides timestamped speech-to-text output.
`operatortextinput`	Text formatting	Cognitive layer	Provides timestamped typed, button, or menu input.
`structuredassemblystate`	Perception layer	Cognitive layer	Integrates scene state, relations, inputs, uncertainty, and evidence.
`reasoningrequest`	Orchestrator Agent	Assembly Reasoning Agent	Packages state, operator input, and knowledge context.
`taskassessmentreport`	Assembly Reasoning Agent	Support Decision Agent	Reports progress, deviations, uncertainty, and rationale.
`supportdecisionpackage`	Support Decision Agent	Orchestrator Agent	Defines proposed support type, target, priority, and rationale.
`cognitivesupportdecision`	Cognitive layer	Guidance layer	Specifies the final support intent for feedback generation.
`guidancecommandpackage`	Guidance layer	Output devices	Contains commands for projector AR, HoloLens, screen, and audio.

Figure 7 represents the toy-train task as a sequential state model. Each step defines the expected component addition and the feedback condition required to advance to the next state. The assembly begins with the combination of the cabin and truck, continues with the boiler and the four wheels, and ends with the roof component. This state model acts as a lightweight procedural knowledge base for the demonstrator: it specifies the expected part, the previous partial assembly, and the condition that should be verified before the system advances to the next cue.

Figure 8 shows the corresponding physical structure of the demonstrator using an exploded CAD representation. The CAD view clarifies the modular composition of the train and the relation between the physical components, the detection labels used by the perception branch, and the expected assembly states. It also provides a basis for defining target regions for projector-based guidance, component metadata for the knowledge layer, and expected spatial relations between parts.

Together, the state model and the exploded CAD view define the procedural and physical references required by the demonstrator. The state model indicates what should happen at each step, while the CAD representation clarifies which components and spatial relations must be detected or guided.

6.3 Experimental setup

After defining the toy-train procedure and component structure, the demonstrator was instantiated in a laboratory workstation to illustrate the perception and guidance elements of the proposed architecture. Figure 9 shows the experimental setup used for the prototype-oriented implementation. The workstation includes three cameras distributed around the assembly workspace, an overhead projector directed toward the work area, and the 3D-printed train components used in the assembly task. Figure 9(a) shows the initial configuration, where the train parts are placed on the workstation before assembly. Figure 9(b) shows the completed train after the sequence has been executed.

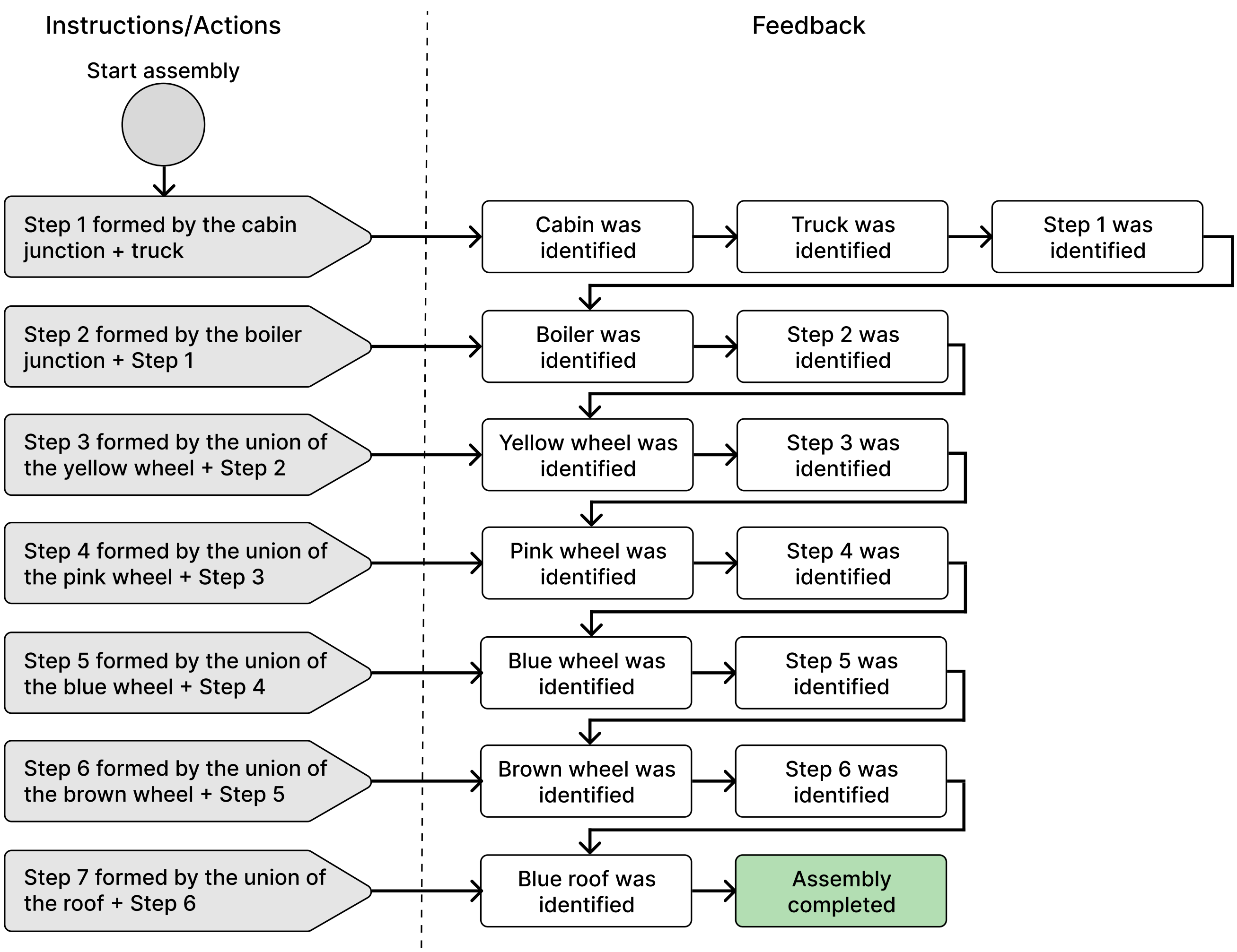


Fig. 7 Toy-train assembly procedure represented as a sequential state model

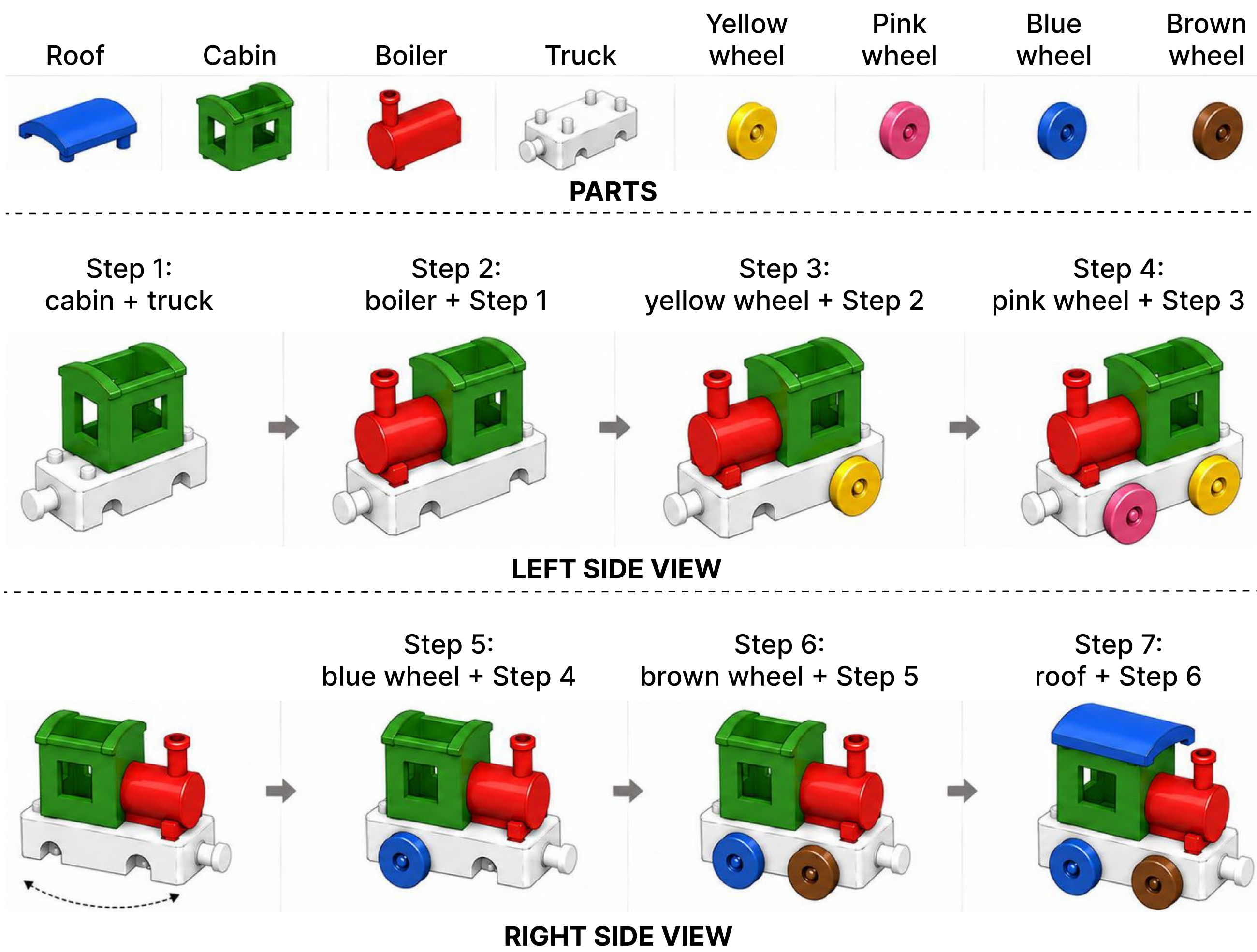
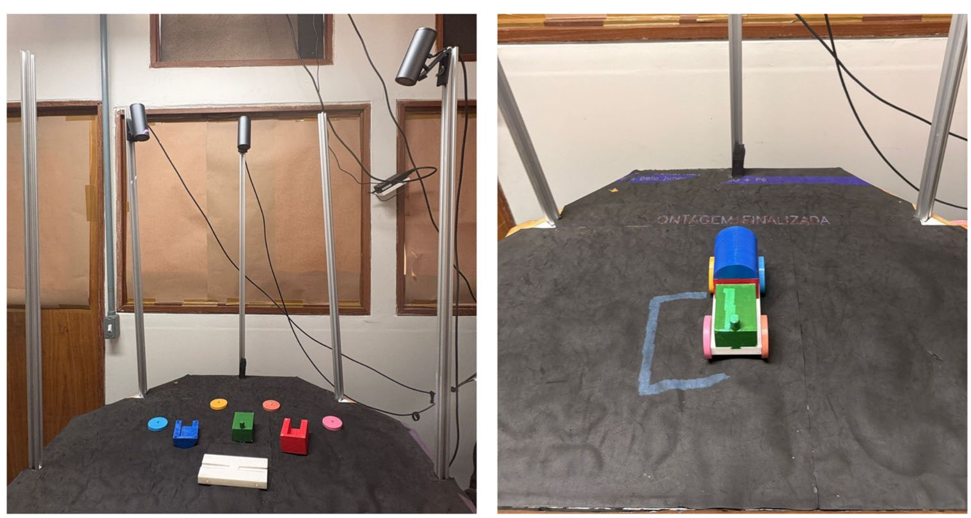


Fig. 8 Exploded CAD representation of the modular toy-train assembly demonstrator

Fig. 9 Experimental setup: (**a**) workstation with cameras, projector, and train components; (**b**) completed train assembly



(a) (b)

The setup was designed to provide both sensing and feedback capabilities. The camera arrangement captures complementary views of the workspace, including the train components, the operator interaction area, and the projected placement regions. The projector provides spatial feedback by displaying cues directly on the assembly surface, such as target regions, component indications, or state-related messages. In this way, the workstation instantiates the physical layer of the proposed architecture and provides the input and output channels required by the perception-cognition-guidance loop.

6.4 Representative perception sequence

Figure 10 shows four representative moments from the toy-train assembly sequence. The selected frames illustrate how the prototype uses projector-based feedback to communicate the current assembly status and the next expected action. In Fig. 10(a), the system provides feedback after identifying the cabin-related state. In Fig. 10(b), the yellow wheel has been identified and the projected cue indicates the next part localization for the pink wheel. Figure 10(c) shows a later step in which the operator is guided to position the roof component. Finally, Fig. 10(d) shows the completed assembly condition, where the final step is identified.

Figure 11 illustrates how one representative scene from the assembly sequence is converted into a perception-layer data contract. The selected scene corresponds to Step 3 of the toy-train procedure, in which the yellow wheel is expected to be joined with the partial assembly obtained after Step 2. The real scene is captured by the camera system and processed by the YOLO11-based object detection and tracking module. The resulting detections are formatted as a `perception_evidence` object. Hand tracking was not implemented in this demonstration; therefore, the example focuses on YOLO11-based object detection, while hand-pose and gesture information remain future inputs to be implemented from the proposed architecture.

The `perception_evidence` contract preserves the evidence required by downstream layers without making a final decision about the assembly state. In this example, the contract includes the detected partial train and the yellow

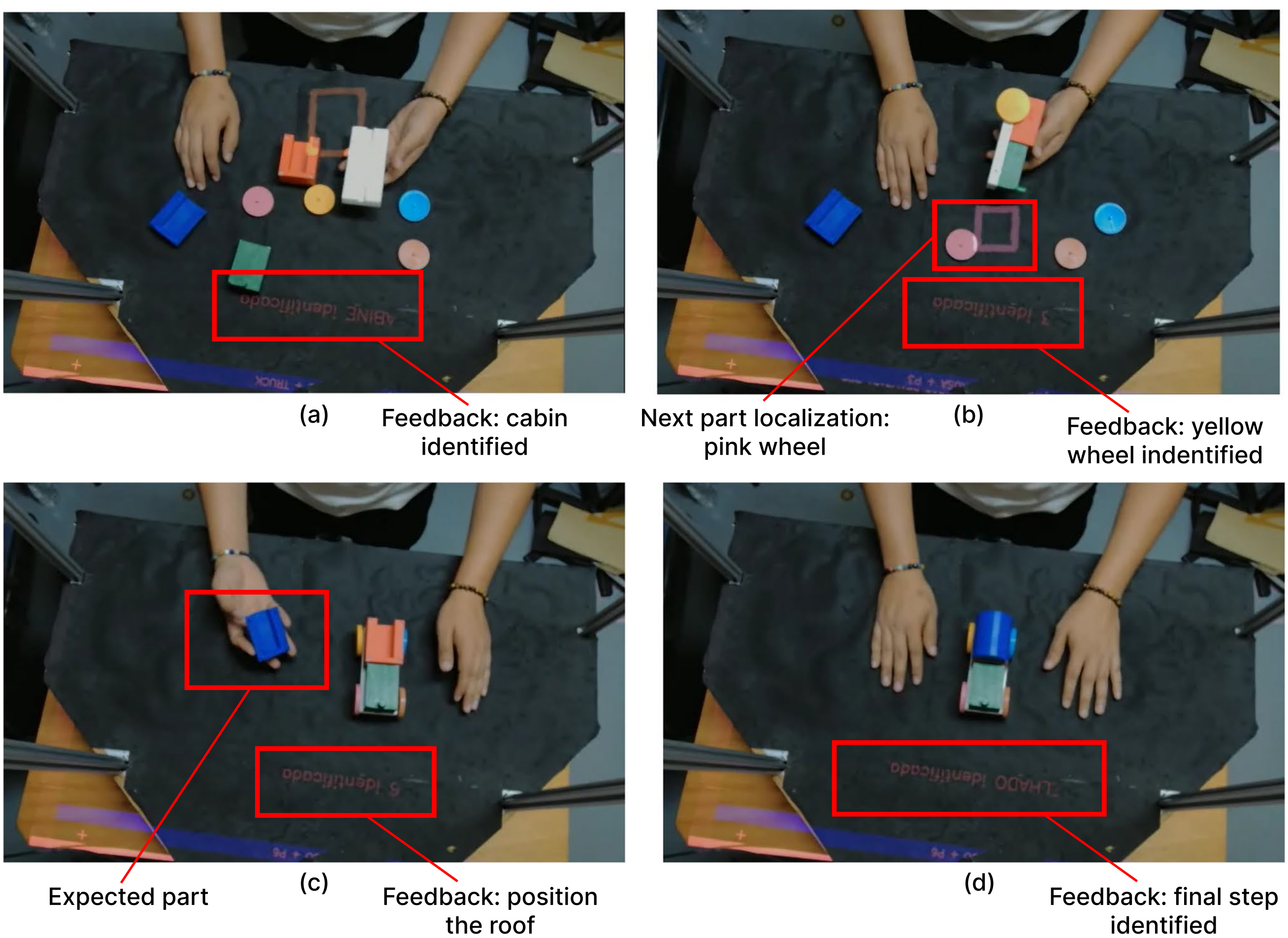


Fig. 10 Representative sequence in the toy-train assembly demonstrator: (**a**) cabin-related state feedback; (**b**) yellow wheel identification and next part localization; (**c**) roof positioning guidance; and (**d**) final assembly-state feedback

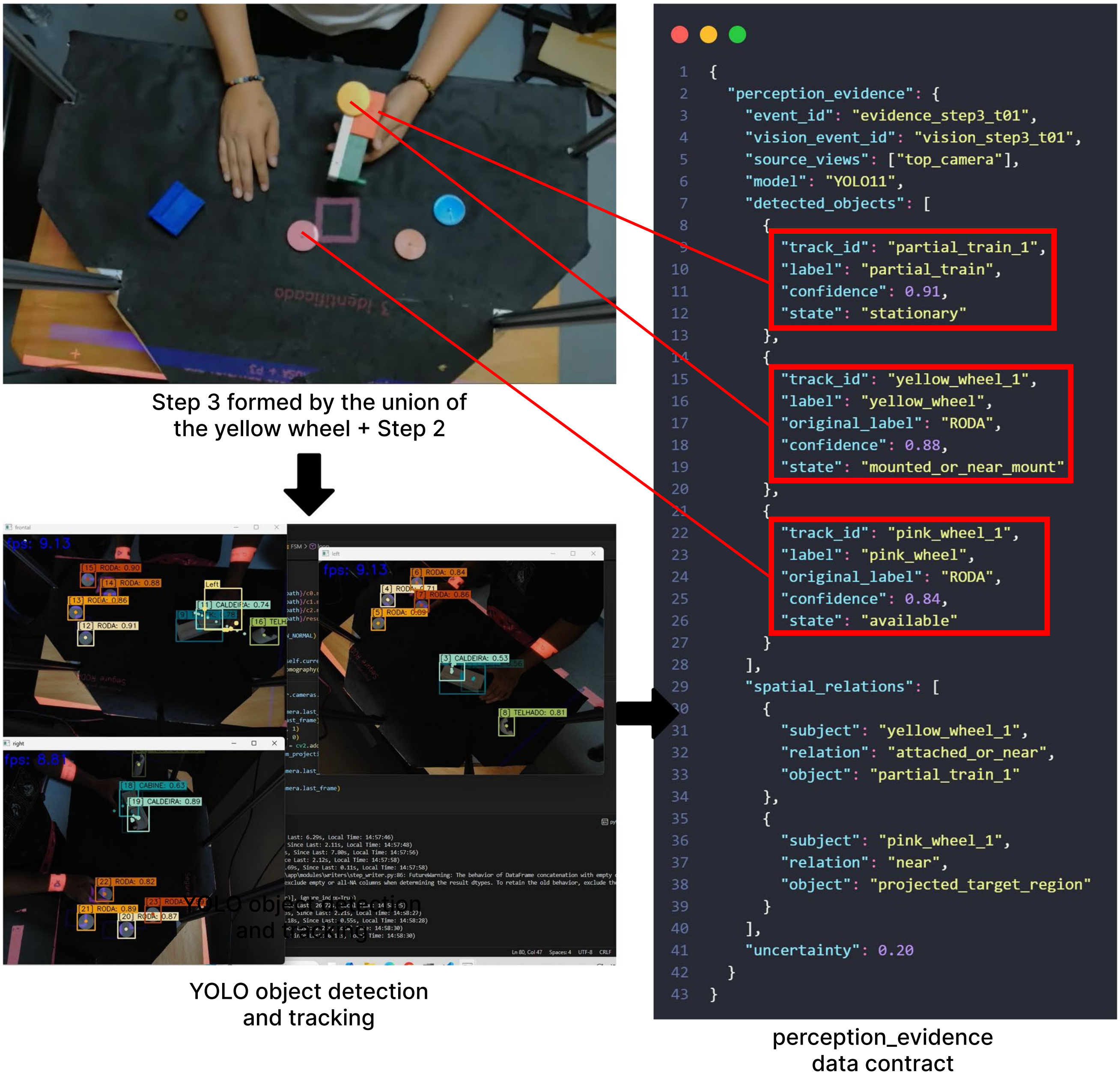


Fig. 11 Representative perception evidence contract for Step 3, where the yellow wheel is detected as the expected component and the pink wheel remains available for the next step

wheel, together with their identifiers, normalized labels, confidence values, and object states. It may also include the pink wheel as an available object in the scene, since it corresponds to the next component in the procedural sequence. Spatial relations describe the yellow wheel as attached to, or near, the partial train, while the pink wheel may be associated with the projected target region for the subsequent step.

This example is important because it separates visual evidence from procedural interpretation. The perception layer does not directly decide that Step 3 is complete or that Step 4 should begin. Instead, it provides structured evidence that can be compared with the state model. If the cognitive layer determines that the yellow wheel satisfies the Step 3 condition, the system can advance the guidance logic toward Step 4, where the pink wheel becomes the next expected component. Therefore, the `perception_evidence` object acts as the traceable input for constructing a `structured_assembly_state` and, later, a `cognitive_support_decision`.

6.5 Contract flow across perception, cognition, and guidance

Figure 12 extends the previous perception example by showing how the evidence generated for Step 3 is propagated through the remaining data contracts of the architecture. The `perception_evidence` object acts as the initial event produced by the perception layer. This evidence is then used to construct a `structured_assembly_state`, where the current status is represented as `yellow_wheel_identified`, the current state is associated with Step 3, and the next expected part is defined as the `pink_wheel`. In parallel, a selected frame can be submitted to a VLM to obtain a `vlm_scene_observation`, which provides a semantic description of the scene and identifies possible evidence limitations, such as attenuated projected text or the need to confirm wheel placement through object localization.

The structured assembly state is then passed to the cognitive layer, where the procedural state model is used to interpret the evidence. In the example, the yellow wheel has been identified, indicating that Step 3 is complete and that the next expected operation is Step 4, corresponding to the placement of the pink wheel. The resulting `cognitive_support_decision` does not directly control the projector; instead, it specifies the support intent, including the support type, target component, message, rationale, confidence, and next expected state. This separation is important because it decouples reasoning from device-specific feedback generation.

Finally, the guidance layer translates the support decision into a `guidance_command_package`. In this case, the command is directed to the projector and specifies a highlighted region associated with the pink-wheel target and the message *Place pink wheel*. The final scene in Fig. 12 illustrates the resulting operator-facing cue. Thus, the example shows how the architecture maintains traceability from visual evidence to state interpretation, cognitive decision-making, and spatial guidance.

Although the contracts are independent of a specific middleware, they can be exchanged between modules through event-based messaging. In the prototype implementation, an MQTT-style communication pattern can be used to decouple the layers: the perception layer publishes `perception_evidence`, the state manager publishes `structured_assembly_state`, the cognitive layer publishes `cognitive_support_decision`, and the guidance layer subscribes to the corresponding decision topic to generate the `guidance_command_package`. This mechanism is not required by the conceptual framework, but it provides a lightweight way to support modularity, asynchronous processing, and traceability between perception, reasoning, and guidance components.

6.6 Preparation and runtime operation modes

Figure 13 complements the layered architecture by distinguishing between preparation mode and runtime support mode. While the layered architecture describes the system's structural organization, the preparation/runtime view describes its operational lifecycle. This distinction is important because the proposed system is not deployed as a single monolithic artifact. It must first be configured, calibrated, trained, and validated at the lab before it can operate online during assembly on the shop floor

Preparation mode corresponds to the offline activities required to build and configure the system components. At the physical level, this includes defining the workstation layout and ergonomics, selecting and calibrating cameras, microphones, AR devices, displays, and other sensors, and preparing tools, fixtures, equipment, tasks, and scenarios. The outputs of this stage include a calibrated setup, workstation configuration, task scenarios, synchronized sensor data, and the physical conditions required for reliable sensing and feedback.

At the perception level, preparation mode includes collecting and annotating visual data, defining label sets and taxonomies, training or fine-tuning object detection models, configuring hand and pose models, and calibrating the multi-sensor setup and data synchronization procedures. The prepared resources include perception models, model parameters, label definitions, and synchronization configurations. These resources allow the runtime system to detect parts, tools, objects, human pose, and assembly-relevant actions in real time.

At the cognitive level, preparation mode includes defining agent roles, configuring prompts, instructions, skills, and tools, building the orchestration workflow, connecting the agents to the knowledge base, and specifying reasoning strategies and decision policies. The outputs include agent configurations, orchestration workflows, reasoning schemas, and support decision rules. This preparation is necessary because the agentic layer should not operate as an unconstrained conversational system; it must reason within the boundaries of the assembly task, procedural knowledge, safety constraints, and human-centered support policies.

At the guidance level, preparation mode includes creating AR assets, 3D models, animations, overlays, instruction templates, voice commands, warnings, feedback rules, UI layouts, and interaction flows. It also includes configuring how guidance is delivered through projector AR, HoloLens MR, screen UI, and audio. The outputs include AR assets, instruction templates, UI layouts, rendering configurations, and feedback policies.

At the knowledge level, preparation mode includes initializing the knowledge base and memory schema, modeling

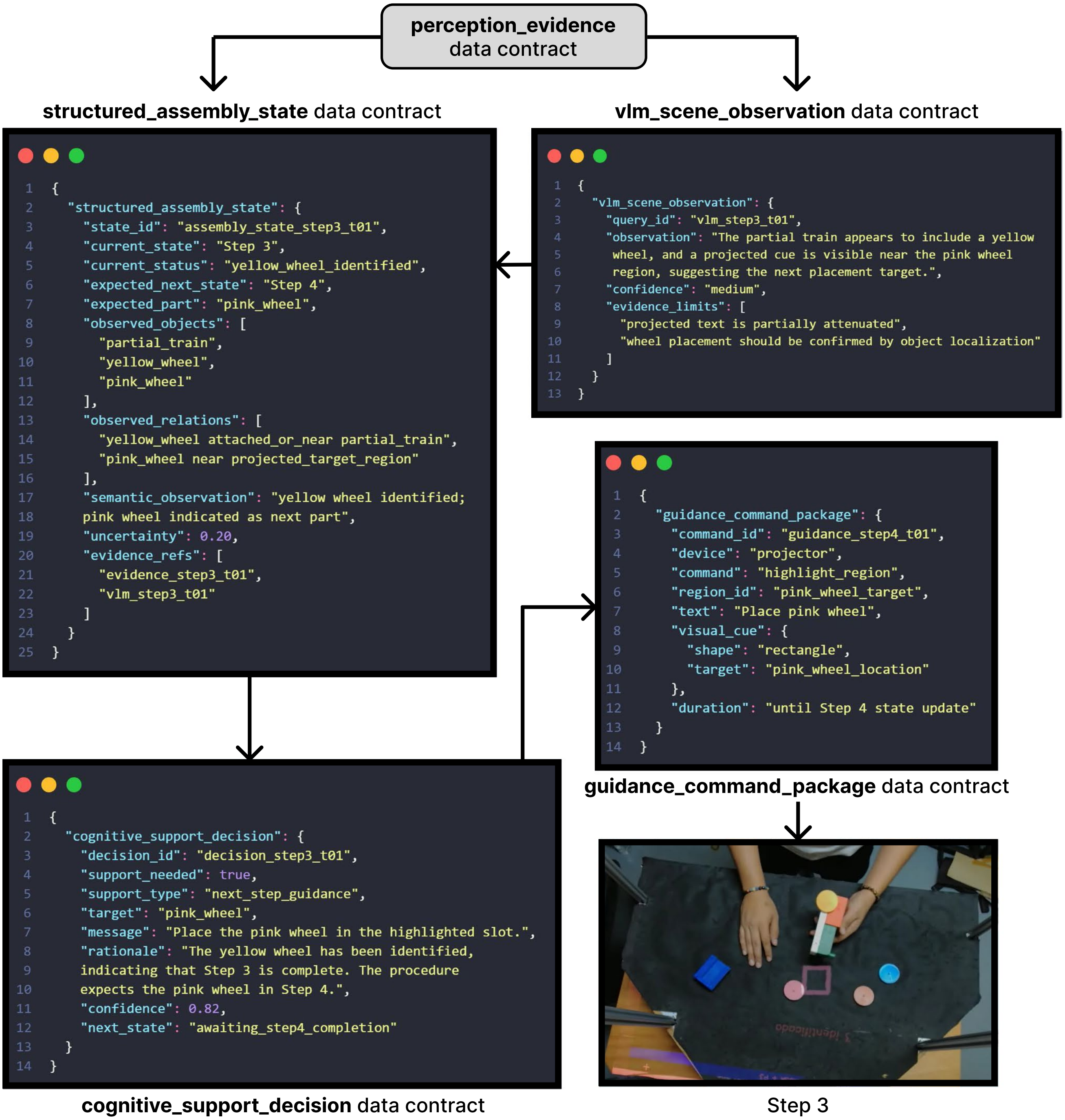


Fig. 12 Representative data-contract flow from Step 3 perception evidence to Step 4 pink-wheel guidance in the toy-train demonstrator

assembly procedures and task sequences, defining constraints, tools, part information, operator profiles, and relevant contextual data. These prepared knowledge resources provide the grounding required for runtime reasoning and guidance.

Runtime support mode corresponds to the system's online operation during assembly. At the physical level, the operator performs the task, manipulates components and tools, and interacts with the workstation. Sen-sors capture real-time video, audio, text, and other signals, while output devices deliver situated guidance. The obtained results include operator actions, environment changes, raw sensor streams, and interaction events.

At the perception level, the system acquires and processes real-time visual and sensor data, detects objects, tracks human pose, estimates actions, and produces a structured

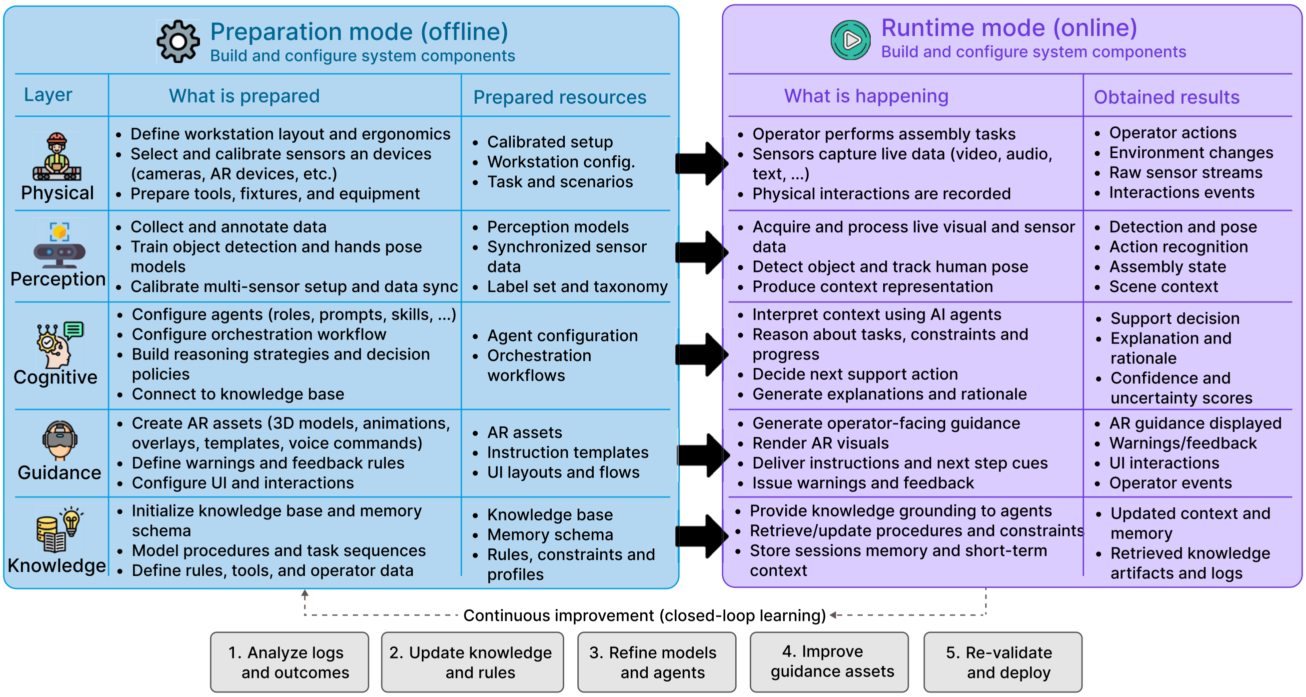


Fig. 13 Preparation and runtime workflow for the cognitive support system in manual industrial assembly

context representation. The outputs include detections, pose information, action recognition, assembly state, and scene context. At the cognitive level, AI agents interpret this context, reason about tasks, constraints, and progress, decide the next support action, and generate explanations and rationale. The outputs include support decisions, confidence and uncertainty information, and reasoning traces. At the guidance level, the system generates operator-facing support, renders AR visuals, delivers instructions and next-step cues, and issues warnings or feedback. The outputs include displayed AR guidance, warnings, UI interactions, and operator events. At the knowledge level, the system retrieves and updates procedures, constraints, session memory, short-term context, and interaction logs.

The two modes are connected by prepared resources. Models, parameters, knowledge bases, AR assets, templates, agent configurations, and system configurations generated offline are used during runtime to provide adaptive support. Conversely, runtime events generate new evidence that can be used to refine the system.

6.7 Continuous improvement and redeployment

The implementation architecture also includes a continuous improvement loop. Runtime operation generates logs, feedback, errors, delays, uncertainty events, operator responses, and support outcomes. These data are analyzed to identify whether problems originate from perception limitations, incomplete knowledge, weak reasoning policies, unsuitable guidance assets, device calibration issues, or interaction design problems.

The improvement loop includes five activities. First, logs and outcomes are analyzed to identify recurring errors, delays, operator difficulties, and unnecessary interventions. Second, knowledge and rules are updated by refining procedures, constraints, task models, and expected states. Third, models and agents are refined by retraining perception models, adjusting prompts, revising agent tools, and updating decision policies. Fourth, guidance assets are improved by modifying AR overlays, instruction templates, warnings, UI flows, and feedback strategies. Finally, the updated configuration is revalidated and redeployed.

This loop is essential because cognitive support systems for manual assembly must adapt not only to technical performance metrics but also to operator behavior, task variability, and usability constraints. A system may fail because a model does not detect a part, because a camera is poorly positioned, because a rule is incomplete, because an agent prompt is ambiguous, or because the displayed guidance is not understandable. Treating implementation as a preparation-runtime improvement cycle makes these dependencies explicit and supports progressive maturation of the system across training, production, ramp-up, research, prototyping, quality support, and maintenance contexts.

7 Discussion and future research directions

The proposed framework and implementation architecture respond to a recurrent limitation identified in the literature: cognitive support in manual assembly is often addressed through isolated technologies rather than through integrated human-centered systems. AR-based solutions are commonly used to present instructions or spatial cues, while AI-based methods are increasingly used for object recognition, error detection, or adaptive information delivery. However, these functions are rarely connected into a closed loop that observes the assembly state, reasons over procedural knowledge, generates context-aware guidance, and learns from operator feedback. The main contribution of the proposed approach is therefore not the use of a single technology, but the integration of perception, agentic reasoning, knowledge grounding, and augmented guidance into a coherent support architecture.

7.1 From visual computation to cognitive support

Recent work on cognition-augmented assembly has shown that computer vision can support human cognition by addressing cognitive demands related to spatiality, memory, knowledge, and decision-making [2]. In practical assembly terms, spatiality concerns the ability to determine where relevant parts, tools, or target locations are; memory concerns the recognition of parts, features, resources, and assembly states; knowledge concerns the interpretation of process context, quality conditions, and potential errors; and decision-making concerns the selection of appropriate assembly actions. This perspective is useful because it clarifies that cognitive support is not only a matter of displaying instructions, but also of helping the operator interpret and act within a changing assembly situation.

The proposed framework extends this logic by connecting visual computation with agentic reasoning and augmented guidance. In this architecture, perception is not limited to object detection or pose estimation; it produces structured evidence about the assembly situation. Cognition is not reduced to classification or rule checking; it interprets the structured state against procedures, constraints, memory, and operator input. Guidance is not only visualization; it is the situated delivery of cues, warnings, explanations, or confirmations according to the current task context. Therefore, the perception-cognition-guidance loop provides a way to move from cognition-augmented visual computation toward integrated cognitive support.

7.2 Assembly task characteristics and application scenarios

The framework can be instantiated across different deployment contexts, including training, production, ramp-up, research, prototyping, quality support, rework, and maintenance. These contexts differ in the type of support required, the acceptable degree of system intervention, and the level of operator expertise. In training, the system can support novice operators by providing step-by-step guidance, explaining task logic, detecting common errors, and progressively reducing assistance as competence increases. In production, it can provide real-time support by monitoring task progress, detecting missing or incorrectly positioned components, and delivering situated cues without forcing the operator to leave the work context.

In ramp-up and new product introduction, the framework can support the stabilization of new assembly processes. At this stage, procedures are often still being refined, operators may not yet be familiar with the product, and recurring difficulties may not be fully understood. By logging operator actions, deviations, support decisions, and feedback, the system can help identify unclear instructions, problematic task sequences, and opportunities for process improvement. In research and prototyping, the architecture can be used as an experimental platform for evaluating human-AI interaction strategies, comparing AR guidance modalities, testing perception models, and studying the role of agentic reasoning in assembly assistance. In quality-oriented support, the framework can contribute to inspection, error prevention, and rework by linking perception evidence with procedural constraints and guidance actions.

The framework can also be interpreted according to assembly task characteristics. For key-part assembly, where large or critical components require accurate positioning and verification, the perception and knowledge layers can support spatial registration, deviation awareness, and quality-oriented guidance. For flexible part assembly, where medium-complexity products involve changing states and task sequences, the structured assembly state and agentic reasoning layer can support contextual interpretation and adaptive guidance. For similar part assembly, where many small or visually similar components must be distinguished, the perception layer can support fine-grained recognition and error prevention, while the guidance layer can reduce operator confusion through spatially situated cues. This task-oriented interpretation helps clarify where the proposed architecture may provide greater value and where simpler assistance systems may be sufficient.

7.3 Implications for assembly-state representation

A central implication of the proposed architecture is that assembly-state representation becomes the main bridge between perception and cognition. In assembly tasks, state representation is particularly challenging because adjacent states may differ only by a single installed, missing, misplaced, or incorrectly oriented component. Simple object detection is therefore insufficient. A useful structured assembly state should encode not only detected objects, but also expected task step, installed-part evidence, missing or redundant components, hand-object relations, object-object relations, uncertainty, evidence references, and possible next states.

This requirement is especially important for agentic reasoning. If the cognitive layer receives only unstructured detections or natural-language descriptions, reasoning becomes difficult to validate and trace. By contrast, a structured assembly state allows the system to compare observed and expected conditions, identify deviations, decide whether support is required, and explain the rationale behind guidance. This also reduces the risk of generative AI producing plausible but invalid recommendations because support decisions are grounded in explicit evidence and procedural constraints.

The structured data contracts proposed in this work should therefore be understood not only as software interfaces, but also as epistemic boundaries between system functions. They define what is known, what is uncertain, what evidence supports the current interpretation, and what information is passed to the next layer. This is particularly relevant in manual assembly, where perception errors, ambiguous visual evidence, and incomplete knowledge can directly affect guidance quality.

7.4 Perception under occlusion, ambiguity, and customization

Manual assembly workspaces are visually complex. Parts may be occluded by hands, tools, fixtures, or other components; visually similar parts may be difficult to distinguish; lighting conditions may vary; and assembly states may change locally as components are progressively installed. These issues are not minor implementation details. They affect whether the system can reliably construct the assembly state required for cognitive support.

For this reason, the perception layer should not attempt to produce definitive interpretations from a single model output. Instead, it should preserve uncertainty and evidence references so that the cognitive layer can reason over incomplete observations, request additional input, or select conservative guidance actions. Multi-camera sensing, hand tracking, VLM-assisted interpretation, and human feedback can all contribute to this process, but only if their outputs are represented in a structured and traceable way.

The architecture also opens the possibility of combining exocentric and egocentric sensing. Exocentric sensing, obtained from fixed cameras observing the workstation, provides stable workspace monitoring and supports process tracking. Egocentric sensing, obtained from wearable or AR devices, captures the operator's visual perspective and may reduce ambiguity in hand-object interactions or occluded regions. Fusing both perspectives could improve assembly-state representation and enable more reliable guidance, particularly in tasks where the operator's viewpoint is central to understanding intent or uncertainty.

Customization introduces another important challenge. In small-batch and high-variety assembly, collecting real images for every product variant can be costly and impractical. Preparation mode should therefore consider synthetic or semi-synthetic data generation from CAD models (like STEP AP 242), 3D product representations, or simulated assembly scenes. Such data can support perception-model training, AR asset preparation, scenario validation, and early testing before runtime deployment. This direction is consistent with the role of the knowledge layer, which can connect product models, task sequences, labels, and guidance assets before the system is used with operators.

7.5 Toward proactive but controlled guidance

The proposed framework supports a shift from reactive assistance to proactive cognitive support. In conventional systems, the operator often has to request information explicitly or follow predefined instructions. In contrast, a perception-cognition-guidance loop can detect uncertainty, deviations, missing components, or operator hesitation and decide whether support is needed. This creates the possibility of proactive guidance, where the system provides assistance before an error occurs or before the operator must stop the task to search for information.

However, proactive support must be carefully constrained. The goal is not continuous unsolicited intervention, but context-aware support when uncertainty, deviation, risk, or operator need is detected. Excessive guidance can increase cognitive load, reduce situational awareness, and undermine operator autonomy. Therefore, the guidance layer should consider not only what information is correct, but also whether it is necessary, when it should be delivered, where it should appear, and which modality is least disruptive.

Agentic reasoning is useful in this context because support decisions often require more than pattern recognition. The system may need to interpret incomplete evidence, compare the current state with procedural knowledge, account for operator input, select between warning, confirmation, explanation, or silence, and justify its recommendation. Nevertheless, the cognitive layer should not behave as an unconstrained conversational system. It must operate within procedural knowledge, safety constraints, human-centered policies, and traceable decision rules. This balance between proactivity and control is essential for responsible cognitive support in real assembly environments.

7.6 Future research directions

Future research should focus on transforming the proposed reference architecture into a validated end-to-end cognitive support system. A first direction is the integration and validation of the complete perception-cognition-guidance loop, including robust assembly-state tracking, uncertainty representation, and reliable mapping between detected objects, procedural steps, support decisions, and guidance targets. Future implementations should also evaluate how perception evidence, VLM-based scene observations, structured assembly states, and guidance commands can be synchronized during real-time assembly execution.

A second direction is the development of safe and traceable agentic reasoning mechanisms. Future work should examine how LLM-based agents can generate support decisions while remaining constrained by validated procedural knowledge, explicit rules, uncertainty thresholds, and human approval. This includes reasoning logs, explanation mechanisms, fallback strategies, and guardrails against invalid, unsafe, or excessive guidance.

A third direction concerns multimodal perception and interaction. Future implementations should evaluate how exocentric cameras, egocentric AR sensing, hand tracking, speech, text, gestures, and operator feedback can be fused into a coherent assembly-state representation. This also requires studying when VLMs should be invoked for semantic interpretation and when faster deterministic models are sufficient.

A fourth direction is adaptive AR guidance. Future studies should compare projected highlights, HoloLens overlays, screen-based instructions, audio prompts, and multimodal feedback in terms of cognitive load, attention, task performance, error reduction, trust, and user acceptance. The goal should not be to maximize information delivery, but to provide the minimum effective support required by the operator and task context.

A fifth direction is preparation-mode automation. Synthetic data generation from CAD models based on STEP AP 242, automatic creation of AR assets, knowledge extraction from process documentation, and reusable data contracts could reduce the engineering effort required to deploy cognitive support systems in high-variety assembly environments.

Finally, future work should evaluate continuous improvement mechanisms. Runtime logs, operator feedback, perception failures, and support outcomes can be used to refine perception models, update knowledge resources, improve prompts, revise guidance assets, and redesign workstation configurations. However, these updates should remain human-reviewed and validated before redeployment in industrial contexts. This is especially important because the purpose of the framework is not to replace human judgment, but to create a traceable and adaptive support system that strengthens human capability in manual assembly.

8 Conclusions

This paper addressed the need for more adaptive and cognitively informed support systems for manual industrial assembly. Although manual assembly remains essential in high-variety, customized, and frequently changing production environments, current assistance systems often remain focused on static work instructions, visual guidance, or isolated perception functions. As a result, operators continue to face cognitive demands related to task variability, information search, memory, error prevention, and situated decision-making.

To examine this problem, the study conducted a systematic search, bibliometric overview, and literature analysis of 129 scientific documents related to cognitive support in manual assembly. The analysis confirmed growing interest in human-centered assembly systems within the Industry 5.0 context. It also showed that Extended Reality technologies, particularly Augmented Reality, dominate the field because of their ability to provide spatially situated instructions and interactive guidance. At the same time, AI-based approaches are increasingly used for object recognition, contextual interpretation, adaptive information delivery, decision-making, and error detection. However, these technologies are still commonly developed as separate capabilities rather than as parts of an integrated cognitive support loop.

Based on these findings, the paper derived design requirements for human-centered cognitive assembly support, emphasizing context-aware assembly-state representation, cognitive load-sensitive information delivery, spatially situated guidance, procedural knowledge grounding, deviation awareness, adaptive support, human control, traceability, modular implementation, multimodal interaction, and explainable reasoning. Together, these requirements indicate that cognitive support should not be understood as one-way instruction delivery, but as a closed human-centered process that observes the assembly state, reasons over procedural knowledge, and provides appropriate guidance while preserving operator autonomy.

Building on these requirements, the paper proposed a conceptual perception-cognition-guidance framework and translated it into a layered implementation architecture. The architecture defines physical, perception, cognitive, guidance, knowledge, and application layers, supported by structured data contracts, preparation/runtime operation modes, and a continuous improvement mechanism. These elements clarify how raw multimodal data can be transformed into perception evidence, structured assembly states, cognitive support decisions, and device-specific guidance commands.

The toy-train assembly demonstrator provides a prototype-oriented instantiation of this logic. It shows how a procedural state model, a physical workstation, camera-based perception, YOLO11-based object detection, projector-based feedback, and contract-based data exchange can connect real assembly events with structured reasoning and operator-facing guidance. Although the demonstrator does not yet constitute a complete industrial validation, it strengthens the proposal by illustrating traceable data flows from physical activity to perception evidence, state interpretation, support decisions, and guidance commands.

Overall, the main contribution of this work is an integrated framework and reference architecture for human-centered cognitive support in manual industrial assembly. Rather than proposing another isolated AR interface or perception module, the paper organizes perception, agentic reasoning, knowledge grounding, and augmented guidance into a coherent support logic aligned with Industry 5.0 principles. Future work should focus on validating the complete perception-cognition-guidance loop in controlled and industrial assembly settings, including its effects on cognitive load, task performance, error reduction, user acceptance, trust, and operator autonomy. Further research is also needed to develop safe agentic reasoning mechanisms, robust multimodal perception, adaptive AR guidance strategies, and human-reviewed continuous improvement processes for real manufacturing environments.

9 Supplementary information

All data generated or analyzed during this study are included in this published article and available at the corresponding author.

Acknowledgements The authors would like to thank the University of Brasilia, MEC (Ministerio de Educação do Brasil) and CAPES (Coordenação de Aperfeiçoamento de Pessoal de Nível Superior).

Author Contributions Not applicable.

Funding The Article Processing Charge (APC) for the publication of this research was funded by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) (ROR identifier: 00x0ma614). This work is supported by CAPES (Coordenação de Aperfeiçoamento de Pessoal de Nível Superior) and the University of Brasilia.

Data Availability All data generated or analyzed during this study are included in this published article and available at the corresponding author.

Materials Availability Not applicable.

Code Availability Not applicable.

Declarations

Conflicts of Interest/Competing Interests The authors declare no competing interests.

Ethics Approval and Consent to Participate Not applicable.

Consent for Publication All authors agree to the publication of this study.

References

1. Dhanda M, Rogers BA, Hall S, Dekoninck E, Dhokia V (2025) Reviewing human-robot collaboration in manufacturing: opportunities and challenges in the context of industry 5.0. Robot Comput-Integr Manuf 93:102937. https://doi.org/10.1016/j.rcim.2024.102937
2. Pang J, Zheng P, Fan J, Liu T (2025) Towards cognition-augmented human-centric assembly: a visual computation perspective. Robot Comput-Integr Manuf 91:102852. https://doi.org/10.1016/j.rcim.2024.102852
3. Xu X, Lu Y, Vogel-Heuser B, Wang L (2021) Industry 4.0 and industry 5.0–inception, conception and perception. J Manuf Syst 61:530–535. https://doi.org/10.1016/j.jmsy.2021.10.006
4. Xu X, Ji T, Zheng P, Wang L (2026) Human-centric manufacturing: re-thinking, re-justifying, and re-envisioning. J Manuf Syst 84:259–268. https://doi.org/10.1016/j.jmsy.2025.12.001
5. Wang L (2022) A futuristic perspective on human-centric assembly. J Manuf Syst 62:199–201. https://doi.org/10.1016/j.jmsy.2021.11.001

6. Eswaran M, Gulivindala AK, Inkulu AK, Bahubalendruni MR (2023) Augmented reality-based guidance in product assembly and maintenance/repair perspective: a state of the art review on challenges and opportunities. Expert Syst Appl 213:118983
7. Dong Z, Wang T (2024) Artificial intelligence driving perception, cognition, decision-making and deduction in energy systems: state-of-the-art and potential directions. Energy Internet 1:27–33. https://doi.org/10.1049/ein2.12010
8. Lee J, Su H (2025) Agentic ai for smart manufacturing. Manuf Lett 46:92–96. https://doi.org/10.1016/j.mfglet.2025.10.013
9. Lampen E, Lehwald J, Pfeiffer T (2020) Virtual humans in ar: evaluation of presentation concepts in an industrial assistance use case. In: Proceedings of the 26th ACM symposium on virtual reality software and technology. https://doi.org/10.1145/3385956.3418974
10. Hoedt S, Claeys A, Landeghem HV, Cottyn J (2016) Evaluation framework for virtual training within mixed-mo del manual assembly. IFAC-PapersOnLine 49:261–266. https://doi.org/10.1016/j.ifacol.2016.07.614
11. Bahubalendruni MR, Biswal BB (2016) A review on assembly sequence generation and its automation. Proc Inst Mech Eng Part C 230:824–838. https://doi.org/10.1177/0954406215584633
12. Michalos G, Makris S, Papakostas N, Mourtzis D, Chryssolouris G (2010) Automotive assembly technologies review: challenges and outlook for a flexible and adaptive approach. CIRP J Manuf Sci Technol 2:81–91. https://doi.org/10.1016/j.cirpj.2009.12.001
13. Pan C (2005) Integrating CAD files and automatic assembly sequence planning. publisherIowa State University
14. Zhang J et al (2022) Projected augmented reality assembly assistance system supporting multi-modal interaction. Int J Adv Manuf Technol 123:1353–1367. https://doi.org/10.1007/s00170-022-10113-6
15. Fu M, Fang W, Gao S, an Yizhou Chen JH (2022) Edge computing-driven scene-aware intelligent augmented reality assembly. Int J Adv Manuf Technol 119:7369–7381. https://doi.org/10.1007/s00170-022-08758-4
16. Kosch T, Funk M, Schmidt A, Chuang LL (2018) Identifying cognitive assistance with mobile electroencephalography: a case study with in-situ projections for manual assembly. Proc ACM Hum-Comput Interact 2. https://doi.org/10.1145/3229093
17. Eswaran M et al (2026) Extended reality (xr) for industrial assembly: a state-of-the-art review toward human-centric industry 5.0. Expert Syst Appl 307:130877. https://doi.org/10.1016/j.eswa.2025.130877
18. Cohen Y, Faccio M, Galizia FG, Mora C, Pilati F (2017) Assembly system configuration through industry 4.0 principles: the expected change in the actual paradigms. IFAC-PapersOnLine 50:14958–14963. https://doi.org/10.1016/j.ifacol.2017.08.2550
19. Zhao S et al (2025) Industrial foundation models (ifms) for intelligent manufacturing: a systematic review. J Manuf Syst 82:420–448. https://doi.org/10.1016/j.jmsy.2025.06.011
20. Zhou Q et al (2026) Towards zero-shot robot tool manipulation in industrial context: a modular vlm framework enhanced by multimodal affordance representation. Robot Comput -Integr Manuf 98:103161. https://doi.org/10.1016/j.rcim.2025.103161
21. Li Y et al (2026) Large language models for manufacturing. J Manuf Syst 86:516–545. https://doi.org/10.1016/j.jmsy.2026.02.014
22. Fan J et al (2025) Vision-language model-based human-robot collaboration for smart manufacturing: a state-of-the-art survey. Front Eng Manag 12:177–200. https://doi.org/10.1007/s42524-025-4136-9
23. Ren Y, Liu Y, Ji T, Xu X (2025) Ai agents and agentic ai-navigating a plethora of concepts for future manufacturing. J Manuf Syst 83:126–133. https://doi.org/10.1016/j.jmsy.2025.08.017
24. Farahani MA, Khan MI, Wuest T (2026) Hybrid agentic ai and multi-agent systems in smart manufacturing. J Manuf Syst 86:612–623. https://doi.org/10.1016/j.jmsy.2026.04.002
25. Moher D, Liberati A, Tetzlaff J, Altman DG, Group P (2010) Preferred reporting items for systematic reviews and meta-analyses: the prisma statement. Int J Surg 8:336–341. https://doi.org/10.1016/j.ijsu.2010.02.007
26. Wang Z et al (2021) M-ar: a visual representation of manual operation precision in ar assembly. Int J Human–Comput Interact 37:1799–1814. https://doi.org/10.1080/10447318.2021.1909278
27. Simonetto M, Peron M, Fragapane G, Sgarbossa F (2021) Digital assembly assistance system in industry 4.0 era: a case study with projected augmented reality 644–651. https://doi.org/10.1007/978-981-33-6318-2_80
28. Agati SS, Bauer RD, Hounsell MdS, Paterno AS (2020) Augmented reality for manual assembly in industry 4.0: Gathering guidelines. In: 2020 22nd Symposium on virtual and augmented reality (SVR), pp 179–188. https://doi.org/10.1109/SVR51698.2020.00039
29. Gollan B, Haslgruebler M, Ferscha A, Heftberger J (2018) Making sense: experiences with multi-sensor fusion in industrial assistance systems. In: Proceedings of the 5th international conference on physiological computing systems, pp 64–74. https://doi.org/10.5220/0007227600640074
30. Van Acker BB et al (2020) Mobile pupillometry in manual assembly: a pilot study exploring the wearability and external validity of a renowned mental workload lab measure. Int J Ind Ergon 75:102891. https://doi.org/10.1016/j.ergon.2019.102891
31. Lampen E et al (2019) Combining simulation and augmented reality methods for enhanced worker assistance in manual assembly. Procedia CIRP 81:588–593. https://doi.org/10.1016/j.procir.2019.03.160
32. Popper KR (1977) The logic of scientific discovery. Syst Zool 26:361
33. Blasing D, Hinrichsen M, Svennd Bornewasser (2020) Reduction of cognitive load in complex assembly systems. Human Interaction, Emerging Technologies and Future Applications II 495–500. https://doi.org/10.1007/978-3-030-44267-5_75
34. Sochor R, Schick TS, Merkel L, Braunreuther S, Reinhart G (2020) Current knowledge management in manual assembly – further development by the analytical hierarchy process, incentive and cognitive assistance systems. In: Proceedings of the conference on production systems and logistics: CPSL 2020. https://doi.org/10.15488/9662
35. Pang J, Zheng P (2023) An mbd-enabled digital twin modeling method for cognition assistance in human-centric smart assembly 1–6. https://doi.org/10.1109/CASE56687.2023.10260573
36. Fang W, Chen L, Han L, Ding J (2025) Context-aware cognitive augmented reality assembly: past, present, and future. J Ind Inf Integr 44:100780. https://doi.org/10.1016/j.jii.2025.100780
37. Caprari G, Katiraee N, Berti N, Battini D (2025) Cognitive ergonomics assessment in manual and collaborative assembly systems summer school francesco turco. Proceedings
38. Pelosi M, Zanchettin AM, Rocco P (2025) Combined hidden-markov model and siamese network approach for assembly operation recognition and error detection. International Journal of Production Research 1–23. https://doi.org/10.1080/00207543.2025.2556484
39. Giridhar MP, Panicker VV (2023) Does cognitive aspects of information and material presentation matter in worker allocation in an assembly line? a case study of a recycling unit in India. Sādhanā 48:23. https://doi.org/10.1007/s12046-023-02078-3
40. Müller R, Vette-Steinkamp M, Hörauf L, Speicher C, Bashir A (2018) Worker centered cognitive assistance for dynamically created repairing jobs in rework area. Procedia CIRP 72:141–146. https://doi.org/10.1016/j.procir.2018.03.137
41. Lampen E, Lehwald J, Pfeiffer T (2020) A context-aware assistance framework for implicit interaction with an augmented human. Virtual, Augmented and Mixed Reality. Industrial and

Everyday Life Applications 91–110. https://doi.org/10.1007/978-3-030-49698-2_7

42. Müller R, Hörauf L, Bashir A (2019) Cognitive assistance systems for dynamic environments. In; 2019 24th IEEE international conference on emerging technologies and factory automation (ETFA), pp 649–656. https://doi.org/10.1109/ETFA.2019.8868986
43. Margherita P, Marcello P (2017) An ergonomics study on manual assembly process re-design in manufacturing firms. https://doi.org/10.3233/978-1-61499-779-5-349
44. Wang B, Zheng L, Wang Y, Wang L, Qi Z (2025) Context-aware ar adaptive information push for product assembly: aligning information load with human cognitive abilities. Adv Eng Inform 64:103086. https://doi.org/10.1016/j.aei.2024.103086
45. Fink K, Ziegler A, Härdtlein C, Berger C, Daub R (2024) Assessing the human competence and individual support level in manual assembly through cognitive assistance systems. Procedia CIRP 126:811–816. https://doi.org/10.1016/j.procir.2024.08.263
46. Wang Z et al (2024) Visual encoding method for mr interface operation process prompts supporting blind area assembly. In: 2024 IEEE 2nd international conference on control, electronics and computer technology (ICCECT), pp 113–117. https://doi.org/10.1109/ICCECT60629.2024.10546058
47. Kosch T, Abdelrahman Y, Funk M, Schmidt A (2017) One size does not fit all: challenges of providing interactive worker assistance in industrial settings. In: Proceedings of the 2017 ACM international joint conference on pervasive and ubiquitous computing and proceedings of the 2017 ACM international symposium on wearable computers, pp 1006–1011. https://doi.org/10.1145/3123024.3124395
48. Li W, Xu A, Wei M, Zuo W, Li R (2024) Deep learning-based augmented reality work instruction assistance system for complex manual assembly. J Manuf Syst 73:307–319. https://doi.org/10.1016/j.jmsy.2024.02.009
49. Pankok C et al (2017) The effects of interruption similarity and complexity on performance in a simulated visual-manual assembly operation. Appl Ergon 59:94–103. https://doi.org/10.1016/j.apergo.2016.08.022
50. Peltokorpi J, Jaber MY (2022) Interference-adjusted power learning curve model with forgetting. Int J Ind Ergon 88:103257. https://doi.org/10.1016/j.ergon.2021.103257
51. Alkan B (2019) An experimental investigation on the relationship between perceived assembly complexity and product design complexity. Int J Interact Des Manuf (IJIDeM) 13:1145–1157. https://doi.org/10.1007/s12008-019-00556-9
52. Solmaz S, Henderickx R, den Bergh JV, Birem M (2024) Capturing the system requirements for adopting industry 5.0 in quality inspection: an evidence-based approach. Procedia CIRP 128:369–374. https://doi.org/10.1016/j.procir.2024.03.016
53. Simmen Y, Eggler T, Legath A, Agotai D, Cords H (2023) Non-overlayed guidance in augmented reality: User study in radiopharmacy. Intelligent Human Computer Interaction 516–526. https://doi.org/10.1007/978-3-031-27199-1_52
54. Research of Task Complexity Decision System for Manual Assembly Tasks Using Fuzzy Cognitive Maps to Construct Bayesian Networks, Vol. Volume 9: Mechanics of Solids, Structures, and Fluids; Micro- and Nano-Systems Engineering and Packaging; Safety Engineering, Risk, and Reliability Analysis; Research Posters of ASME International Mechanical Engineering Congress and Exposition. https://doi.org/10.1115/IMECE2022-93881
55. Abdul Hadi M et al (2022) Towards flexible and cognitive production—addressing the production challenges. Applied Sciences 12. https://www.mdpi.com/2076-3417/12/17/8696
56. Brolin A, Thorvald P, Case K (2017) Experimental study of cognitive aspects affecting human performance in manual assembly. Prod Manuf Res 5:141–163. https://doi.org/10.1080/21693277.2017.1374893
57. Lindblom J, Thorvald P (2017) Manufacturing in the wild - viewing human-based assembly through the lens of distributed cognition. Prod Manuf Res 5:57–80. https://doi.org/10.1080/21693277.2017.1322540
58. Xiao H, Duan Y, Zhang Z, Li M (2018) Detection and estimation of mental fatigue in manual assembly process of complex products. Assem Autom 38:239–247. https://doi.org/10.1108/AA-03-2017-040
59. Agati SS, da S Hounsell M, Paterno AS (2024) Graal—modeling, prototyping and assessing a gamified responsible augmented assembly line system. Int J Adv Manuf Technol 132:2735–2751. https://doi.org/10.1007/s00170-024-13460-8
60. Papetti A, Ciccarelli M, Brunzini A, Germani M (2022) Investigating the application of augmented reality to support wire harness activities 116–124. https://doi.org/10.1007/978-3-030-91234-5_11
61. Yan Y et al (2023) A novel adaptive visualization method based on user intention in ar manual assembly. Int J Adv Manuf Technol 129:4705–4730. https://doi.org/10.1007/s00170-023-12557-w
62. Pokorni B, Popescu D, Constantinescu C (2022) Design of cognitive assistance systems in manual assembly based on quality function deployment. Applied Sciences 12. https://www.mdpi.com/2076-3417/12/8/3887
63. Kolbeinsson A, Lindblom J, Thorvald P (2017) Missing mediated interruptions in manual assembly: critical aspects of breakpoint selection. Appl Ergon 61:90–101. https://doi.org/10.1016/j.apergo.2017.01.010
64. Lindblom J, Gündert J (2017) Managing mediated interruptions in manufacturing: selected strategies used for coping with cognitive load 389–403. https://doi.org/10.1007/978-3-319-41691-5_33
65. Wollter Bergman M, Berlin C, Babapour Chafi M, Falck A-C, Örtengren R (2021) Cognitive ergonomics of assembly work from a job demands–resources perspective: three qualitative case studies. International Journal of Environmental Research and Public Health 18. https://www.mdpi.com/1660-4601/18/23/12282
66. Kuipers N, Kolbeinsson A, Thorvald P (2021) Appropriate assembly instruction modes: factors to consider. https://doi.org/10.3233/ATDE210007
67. Torres Y, Nadeau S, Landau K (2021) Evaluation of fatigue and workload among workers conducting complex manual assembly in manufacturing. IISE Trans Occup Ergon Human Factors 9:49–63. https://doi.org/10.1080/24725838.2021.1997835
68. Bläsing D, Bornewasser M (2021) Influence of increasing task complexity and use of informational assistance systems on mental workload. Brain Sciences 11. https://www.mdpi.com/2076-3425/11/1/102
69. Acker BBV et al (2021) Development and validation of a behavioural video coding scheme for detecting mental workload in manual assembly. Ergonomics 64:78–102. https://doi.org/10.1080/00140139.2020.1811400
70. Merkel L, Berger C, Braunreuther S, Reinhart G (2019) Determination of cognitive assistance functions for manual assembly systems 198–207
71. Hinrichsen S, Nikolenko A, Beckmann N, Meyer F (2022) Development of a new type of carousel-based compacted work system for mixed-model assembly in mechanical engineering 0087–0090. https://doi.org/10.1109/IEEM55944.2022.9989731
72. Gan ZL, Musa SN, Yap HJ, Liu C, Xu Y (2022) A conceptual framework of adaptive ar-based digital guidance system for high-mix low-volume manual assembly. In: 2022 27th International conference on automation and computing (ICAC), pp 1–5. https://doi.org/10.1109/ICAC55051.2022.9911112
73. Kelm B, Haas PH, Jochum S, Margies L, Müller R (2025) Enhancing assembly instruction generation for cognitive assistance

systems with large language models. Procedia CIRP 134:7–12. https://doi.org/10.1016/j.procir.2025.03.010

74. Gan ZL, Musa SN, Yap HJ (2025) Machine learning on assembly guidance system ontology for manufacturing assembly. Proceedings of the Institution of Mechanical Engineers, Part B: Journal of Engineering Manufacture. https://doi.org/10.1177/09544054251395468
75. Freydank E, Kießling N, Seyfried L, Cencic MR (2026) Transforming industrial training: A comparative study of volumetric video in mixed reality and paper-based instructions 456–466. https://doi.org/10.1007/978-3-031-97763-3_33
76. Sridhar EP et al (2025) Enhancing manual assembly: a comparative study of mixed reality and paper-based instructions assessing user performance and cognitive load. In: IISE annual conference. Proceedings. https://doi.org/10.21872/2025IISE_5071
77. Petzoldt C, Keiser D, Schöbel N, Freitag M (2022) Planung von assistenzsystemen für die industrielle montage. Zeitschrift für wirtschaftlichen Fabrikbetrieb 117:157–163
78. Fink K, Riemensperger T, Brugger M, Berger J, Braunreuther S (2021) Konfigurator für kognitive assistenzsysteme/configuration of cognitive assistance systems. wt Werkstattstechnik Online 111:770–774. https://doi.org/10.37544/1436-4980-2021-10-116
79. Büttner S, Funk M, Sand O, Röcker C (2016) Using head-mounted displays and in-situ projection for assistive systems: a comparison. In: Proceedings of the 9th ACM international conference on PErvasive technologies related to assistive environments. https://doi.org/10.1145/2910674.2910679
80. Giridhar MP, Lasin, Panicker VV (2020) Experimental analysis of cognitive issues impacting manual assembly task. In: 2020 International Conference on System, Computation, Automation and Networking (ICSCAN), pp 1–6. https://doi.org/10.1109/ICSCAN49426.2020.9262305
81. Berlin C, Bergman MW, Chafi MB, Falck A-C, Örtengren R (2021) A systemic overview of factors affecting the cognitive performance of industrial manual assembly workers 371–381. https://doi.org/10.1007/978-3-030-74608-7_47
82. Parmentier DD, Acker BBV, Detand J, Saldien J (2020) Design for assembly meaning: a framework for designers to design products that support operator cognition during the assembly process. Cogn Technol Work 22:615–632. https://doi.org/10.1007/s10111-019-00588-x
83. Peixe LC, Agati S, Hounsell MDS (2024) Manual assembly augmented reality systems implementation: a systematic literature mapping. In: Proceedings of the 25th symposium on virtual and augmented reality, pp 17–25. https://doi.org/10.1145/3625008.3625011
84. Giridhar MP, Panicker VV (2025) Towards measuring cognitive load through personalised models in a medium-scale enterprise: a worker allocation approach. Theor Issues Ergon Sci 26:304–331. https://doi.org/10.1080/1463922X.2024.2438030
85. Renu RS, Righter J, Lytch J (2022) Developing requirements for a manufacturing training platform: a three-pronged approach. Volume 2: 42nd Computers and Information in Engineering Conference (CIE)
86. Hasan N, Alkan B (2025) Gest-sar: a gesture-controlled spatial ar system for interactive manual assembly guidance with real-time operational feedback. Machines 13. https://doi.org/10.3390/machines13080658
87. Oestreich H, Töniges T, Wojtynek M, Wrede S (2019) Interactive learning of assembly processes using digital assistance. Proc Manuf 31:14–19. https://doi.org/10.1016/j.promfg.2019.03.003y
88. Heinz-Jakobs M, Große-Coosmann A, Röcker C (2022) Promoting inclusive work with digital assistance systems: Experiences of cognitively disabled workers with in-situ assembly support 377–384. https://doi.org/10.1109/GHTC55712.2022.9910994
89. Claeys A et al (2022) Methodology to integrate ergonomics information in contextualized digital work instructions. Procedia CIRP 106:168–173. https://www.sciencedirect.com/science/article/pii/S2212827122001743
90. Aehnelt M, Urban B (2015) The knowledge gap: providing situation-aware information assistance on the shop floor 232–243. https://doi.org/10.1007/978-3-319-20895-4_22
91. Müller R, Hörauf L, Speicher C, Bashir A (2019) Situational cognitive assistance system in rework area. Proc Manuf 38:884–891. https://doi.org/10.1016/j.promfg.2020.01.170
92. Bleser G et al (2015) Cognitive learning, monitoring and assistance of industrial workflows using egocentric sensor networks. PLoS ONE 10:e0127769. https://doi.org/10.1371/journal.pone.0127769
93. Strenge B, Schack T (2021) Empirical relationships between algorithmic sda-m-based memory assessments and human errors in manual assembly tasks. Sci Rep 11:9473. https://doi.org/10.1038/s41598-021-88921-1
94. Riedel A et al (2021) A deep learning-based worker assistance system for error prevention: case study in a real-world manual assembly. Adv Prod Eng Manag 16:393–404. https://doi.org/10.14743/apem2021.4.408
95. Shinde PS et al (2025) Transforming msme assembly operations: smart manual assembly table for improved productivity. Int J Basic Appl Sci 14:40–51. https://doi.org/10.14419/am35a906
96. Qeshmy DE, Makdisi J, Ribeiro da Silva EHD, Angelis J (2019) Managing human errors: augmented reality systems as a tool in the quality journey. Proc Manuf 28:24–30. https://doi.org/10.1016/j.promfg.2018.12.005
97. Riedel A et al (2023) Evaluating augmented reality, deep learning and paper-based assistance systems in industrial manual assembly. In: Advances in production management systems. production management systems for responsible manufacturing, service, and logistics futures, pp 417–431. https://doi.org/10.1007/978-3-031-43662-8_30
98. Zigart T, Schlund S (2022) 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. https://doi.org/10.1109/ISMAR-Adjunct57072.2022.00022
99. Sochor R, Kraus L, Merkel L, Braunreuther S, Reinhart G (2019) Approach to increase worker acceptance of cognitive assistance systems in manual assembly. Procedia CIRP 81:926–931. https://doi.org/10.1016/j.procir.2019.03.229
100. Jeffri NFS, Rambli DRA (2020) Guidelines for the interface design of ar systems for manual assembly. In: Proceedings of the 2020 4th international conference on virtual and augmented reality simulations, pp 70–77. https://doi.org/10.1145/3385378.3385389
101. Neumann A et al (2020) Avikom: towards a mobile audiovisual cognitive assistance system for modern manufacturing and logistics. In: Proceedings of the 13th ACM international conference on PErvasive technologies related to assistive environments. https://doi.org/10.1145/3389189.3389191
102. Simões B, Amicis RD, Barandiaran I, Posada J (2019) Cross reality to enhance worker cognition in industrial assembly operations. Int J Adv Manuf Technol 105:3965–3978. https://doi.org/10.1007/s00170-019-03939-0
103. Pratticò FG, Di Cosmo D, Piviotti M, La Rosa G, Lamberti F (2024) Effectiveness of computer-based e-learning and virtual reality training system for the preparatory phase of a training on maintenance procedure 1–6. https://doi.org/10.1109/ICCE59016.2024.10444354
104. Guo Z et al (2022) An evaluation method using virtual reality to optimize ergonomic design in manual assembly and maintenance

scenarios. Int J Adv Manuf Technol 121:5049–5065. https://doi.org/10.1007/s00170-022-09657-4
105. Ulmer J, Braun S, Cheng C-T, Dowey S, Wollert J (2023) A human factors-aware assistance system in manufacturing based on gamification and hardware modularisation. Int J Prod Res 61:7760–7775. https://doi.org/10.1080/00207543.2023.2166140
106. Meara MO, Cheng X, Eden J, Ivanova E, Burdet E (2023) A third eye to augment environment perception 1239–1244. https://doi.org/10.1109/ROMAN57019.2023.10309628
107. Trendov A, Rizov T, Jovanoski B, Trendova KM (2026) Enhancing industrial workflows with augmented reality (ar): Ar-based poka-yoke visual assembly guide for a smart learning factory 157–172. https://doi.org/10.1007/978-3-031-97772-5_11
108. Precup S-A, Pirvu B-C, Gellert A, Zamfirescu C-B (2025) Cognitive control units in smart manufacturing: insights from a soar-based assistance system. IEEE Access 13:196167–196180. https://doi.org/10.1109/ACCESS.2025.3633609
109. Fang W, Teng Z, Zhang Q, Wu Z (2024) A natural bare-hand interface-enabled interactive ar assembly guidance. Int J Adv Manuf Technol 133:3193–3207. https://doi.org/10.1007/s00170-024-13922-z
110. Marino E, Barbieri L, Bruno F, Muzzupappa M (2025) A novel ar recoloring technique to enhance operator performance on inspection tasks in industry 4.0 environments. IEEE Trans Visual Comput Graphics 31:4605–4618. https://doi.org/10.1109/TVCG.2024.3410537
111. Wang Z et al (2024) Visual encoding method for mr interface operation process prompts supporting blind area assembly 113–117. https://doi.org/10.1109/ICCECT60629.2024.10546058
112. Li J et al (2026) An adaptive ar guidance interface layout optimization approach for human-centered assembly training systems. Adv Eng Inform 69:103975. https://doi.org/10.1016/j.aei.2025.103975
113. Wang Z et al (2022) A comprehensive review of augmented reality-based instruction in manual assembly, training and repair. Robot Comput-Integr Manuf 78:102407. https://doi.org/10.1016/j.rcim.2022.102407
114. Singh AK (2025) Vr/ar in ergonomics and workspace design: a dual-perspective analysis of applications and implications. Appl Ergon 129:104612. https://doi.org/10.1016/j.apergo.2025.104612
115. Drouot M, Le Bigot N, Bricard E, de Bougrenet J-L, Nourrit V (2022) Augmented reality on industrial assembly line: impact on effectiveness and mental workload. Appl Ergon 103:103793. https://doi.org/10.1016/j.apergo.2022.103793
116. Sopidis G et al (2022) Micro-activity recognition in industrial assembly process with imu data and deep learning. In: Proceedings of the 15th international conference on PErvasive technologies related to assistive environments, pp 103–112. https://doi.org/10.1145/3529190.3529204
117. Nakayama K, Onari H (2024) Analysis of takt time extension in assembly lines with multiple elemental works allocated to a process 581–589. https://doi.org/10.1007/978-981-97-0194-0_57
118. Sochor R, Merkel L, Braunreuther S, Reinhart G, Greiter F (2020) Monetäre bewertung eines anreizsystems/economic feasibility study of an incentive system for motivating manual assembly employees. wt Werkstattstechnik online 110:125–129. https://doi.org/10.37544/1436-4980-2020-03-41
119. Johansson P et al (2018) Assessment based information needs in manual assembly. DEStech Transactions on Engineering and Technology Research. https://doi.org/10.12783/dtetr/icpr2017/17637
120. Wang X, Ong S, Nee A (2016) Multi-modal augmented-reality assembly guidance based on bare-hand interface. Adv Eng Inform 30:406–421. https://doi.org/10.1016/j.aei.2016.05.004
121. Keiser D, Petzoldt C, Walura V, Leimbrink S, Freitag M (2023) Concept and integration of knowledge management in assembly assistance systems. Procedia CIRP 118:940–945. https://doi.org/10.1016/j.procir.2023.06.162
122. Ganlin Z, Pingfa F, Jianfu Z, Dingwen Y, Zhijun W (2021) Information integration and instruction authoring of augmented assembly systems. Int J Intell Syst 36:5028–5050. https://doi.org/10.1002/int.22501
123. Rupprecht P, Kueffner-McCauley H, Trimmel M, Schlund S (2021) Adaptive spatial augmented reality for industrial site assembly. Procedia CIRP 104:405–410. https://doi.org/10.1016/j.procir.2021.11.068
124. Funk M, Hartwig M, Wischniewski S (2019) Evaluation of assistance systems for manual assembly work 794–798. https://doi.org/10.1109/SII.2019.8700420