Robotics and Computer–Integrated Manufacturing 103 (2027) 103406

Contents lists available at ScienceDirect

Robotics and Computer-Integrated Manufacturing

journal homepage: www.elsevier.com/locate/rcim

Review

Embodied AI in industrial operations: A systematic literature review

Julian Koch *,[1], Philipp Prünte [1], Keno Moenck [1], Jonathan Determann [1], Martin Gomse, Thorsten Schüppstuhl

Institute of Aircraft Production Technology, Hamburg University of Technology, Denickestraße 17, Hamburg, 21073, Germany

A R T I C L E I N F O

Keywords:
Embodied AI
Structured literature review
Industrial robotics
Foundation models
Human–robot collaboration
Manufacturing

A B S T R A C T

Embodied Artificial Intelligence (EAI) is increasingly seen as a promising route toward flexible automation, also in industrial operations. Yet the term is used inconsistently, and existing surveys often address isolated components rather than the combined role of embodiment, learning, perception, action, data, and industrial deployment constraints. This paper presents a PRISMA-guided systematic literature review of recent EAI research in industrial operations. We first consolidate the terminology of EAI and define four guiding characteristics: embodiment, learning, perception-action coupling, and situated intelligence. Based on this framework, we introduce a manipulation-mobility matrix for classifying physical embodiments and analyze the reviewed literature across industrial applications, research focuses, hardware forms, input and output modalities, learning architectures, datasets, simulation environments, and training paradigms. The results show that current research is dominated by assembly and manufacturing scenarios, stationary robotic arms, vision- and language-based interfaces, and modular foundation model-centered architectures. However, tightly coupled sensorimotor learning, tactile and force feedback, situated intelligence, and validated real-world deployment remain limited. This review clarifies the current state of industrial EAI and identifies the key gaps that must be closed to enable scalable deployment in industrial operations.

Contents

* Corresponding author.
E-mail addresses: julian.koch@tuhh.de (J. Koch), philipp.pruente@tuhh.de (P. Prünte), keno.moenck@tuhh.de (K. Moenck), jonathan.determann@tuhh.de (J. Determann), martin.gomse@tuhh.de (M. Gomse), schueppstuhl@tuhh.de (T. Schüppstuhl).
[1] Equal contribution.

https://doi.org/10.1016/j.rcim.2026.103406
Received 15 June 2026; Received in revised form 8 August 2026; Accepted 11 August 2026
Available online 27 August 2026

1. Introduction

Artificial Intelligence (AI)-driven technologies are omnipresent beyond our everyday life. This surge was initially driven largely by the success of natural language chat systems based on Large Language Models (LLMs), especially ChatGPT in 2022, which dominated the discourse [1]. Over the past two years, however, this interest has expanded significantly and is no longer limited to disembodied AI. Instead, Embodied AI (EAI) is increasingly taking center stage, with a particular focus on humanoid robotics [2]. While LLMs primarily function as chatbots and virtual assistants for end users [3,4], examples of EAI are most prominently found in the form of mobile service robots for domestic assistance, such as autonomous cleaning systems [5] and research platforms performing household chores like autonomously tidying up cluttered rooms [6] and folding laundry [7].

In the context of EAI, current discourse highlights a potential paradigm shift, envisioning the automation or fundamental transformation of manual tasks in production environments [8]. While actual, widespread industrial use cases are still rare at present, some early pilot projects already demonstrate immense potential. Promising examples include the deployment of humanoid robots at the BMW plants in Spartanburg and Leipzig [9] and the technological developments in industrial robotics by Hyundai in collaboration with Boston Dynamics [10]. These examples show that manufacturing companies are increasingly exploring EAI to automate tasks previously deemed economically unfeasible. The core novelty of EAI lies in its ability to learn and adapt to complex operations without the extensive effort required by prevailing methods, e.g., traditional programming. By offering this flexibility, these systems could potentially help mitigate persistent labor shortages, e.g., in warehousing, logistics, and highly variable production environments.

The increasing recognition of this industrial potential is also reflected in the academic literature. As illustrated in Fig. 1, the number of both original research articles and review papers dedicated to EAI has surged between 2020 and 2025. This growth underscores the transition of EAI from a theoretical concept to a primary focus of applied robotics research. This momentum is particularly visible in human–robot collaboration for manufacturing, where recent work targets intention perception, imitation-based skill transfer, and collaborative assembly [11,12]. Assessing the viability of EAI in manufacturing requires analyzing five core components. First, the specific industrial applications impose strict precision and safety constraints. Second, the physical embodiment dictates the system's capabilities. Third, the chosen learning strategies determine how the system acquires skills. Fourth, specialized datasets and simulations are required to bridge domain gaps such as the sim-to-real gap. Finally, diverse input modalities are essential to map environmental perception to reliable execution.

Table 1 outlines the thematic scope of recent review papers, revealing a strong focus on isolated EAI components rather than broader application in industrial operations. For instance, several recent works provide overviews of EAI frameworks or object-centric manipulation, but do so without considering the strict operational constraints of industrial environments [13–15]. Other reviews concentrate heavily on the role of foundation models in robotics, yet they often abstract away from the physical realities of diverse embodiments or domain-specific training pipelines [16,17]. Even when industrial use cases are briefly acknowledged, such as in the multi-domain taxonomy of AI agents by Piccialli et al. [18], the analysis remains highly generalized. Consequently, the critical intersection of multimodality, datasets, learning paradigms, and morphological hardware required for industrial operations has yet to be systematically synthesized.

To address the identified gaps in the existing literature reviews, this paper focuses specifically on the industrial context. Unlike previous surveys, this review systematically examines all core technological components of EAI. This entails a detailed analysis of physical embodiments and input modalities alongside the underlying mechanisms for perception, learning, and action. By synthesizing these elements, this paper provides a current overview of the state of EAI for practitioners and researchers. Specifically, the main contributions of this work are:

- **Terminology and Architecture:** We explicitly define the core characteristics of industrial EAI and detail its structural components.
- **Manipulation-Mobility Matrix:** We present a taxonomy to classify EAI hardware based on dexterity and locomotion capabilities of the embodiment.
- **Systematic Literature Review (SLR):** We conduct a PRISMA-guided review to assess the state of EAI in industrial operations from manufacturing, assembly, and logistics to Maintenance, Repair, and Overhaul (MRO). The results of this SLR can be divided into two main categories:
  - **Current Landscape of Industrial EAI:** We map the distribution of current research across industrial applications, research focuses, and physical embodiments, while evaluating the application of the EAI terminology within the field.
  - **Challenges:** By analyzing perception-action coupling, the integration of diverse sensory modalities and outputs, and prevailing training paradigms, we identify the primary technological challenge (e.g., the over-reliance on decoupled foundation models) preventing scalable industrial deployment.

The remainder of this work is structured as follows: Section 2 establishes the terminology and physical taxonomy of EAI. Section 3 details the PRISMA-guided review methodology. The findings of our SLR are presented in Section 4. Section 5 synthesizes the core insights, followed by final remarks on challenges in Section 6.

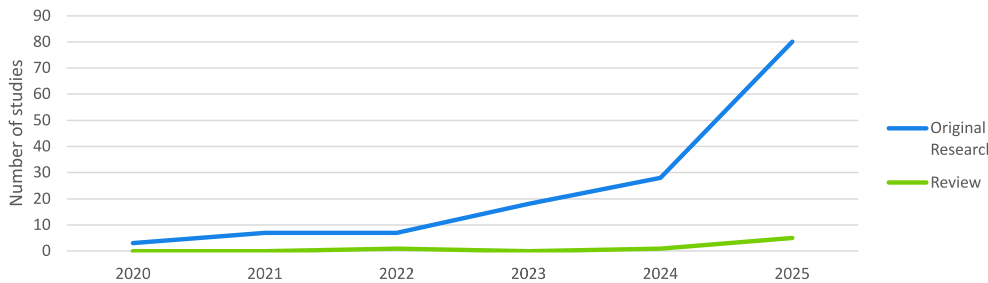

Fig. 1. Number of original research and review papers from 2020 to 2025 addressing EAI. The list of original papers is provided in Appendix A.2. The included original papers discuss EAI in an industrial context. *Note: This graph excludes 8 original papers published in 2025 but listed for 2026.*

Table 1
Recent surveys/reviews in EAI (2020–2025).

Authors	Year	Title	Time span	Goal	Industrial App.	Embodiment	Learning	Datasets	Input Mod.
H. Liu et al.	2025	Embodied Intelligence: A Synergy of Morphology, Action, Perception and Learning [13]	ca. 1990–2025	Propose a unified EAI Framework (M-A-P-L) and organize work based on this structure, as well as reviewing the development of EAI.	✗	✓	RL, F	✗	V, L, O, MM
Y. Zheng et al.	2025	A Survey of Embodied Learning for Object-centric Robotic Manipulation [14]	2021–2025	Systematic map of embodied learning for object-centric manipulation from perception to policy to task execution, plus datasets, metrics, applications, and open problems.	(P)	✗	RL, IL, F	✓	V, T, (MM)
Y. Liu et al.	2025	Aligning Cyber Space With Physical World: A Comprehensive Survey on Embodied AI [15]	2007–2025	Overview of EAI across robots, simulators, and four core targets (embodied perception, interaction, agents, sim-to-real).	✗	✓	F, (RL, IL)	✓	V, L, MM
F. Picialli et al.	2025	AgentAI: A comprehensive survey on autonomous agents in distributed AI for Industry 4.0 [18]	2014–2025	Multi-domain taxonomy of agentic AI for Industry 4.0, unify non-autonomous and fully autonomous agents, review of core technology (LLMs/VLMs, RL, Transformer), path forward to Industry 5.0.	P	✗	RL, F, O	✗	V, L, O
J. Duan et al.	2022	A Survey of Embodied AI: From Simulators to Research Tasks [19]	2016–2020	Survey of EAI for exploration of 3D space (visual exploration, visual navigation, embodied question answering) and benchmark of simulators.	✗	(✓)	RL	✓	V, L, O
Z. Xu et al.	2024	A Survey on Robotics with Foundation Models: toward Embodied AI [17]	2021–2024	Provides a comprehensive, up-to-date overview of how foundation models are used in robotics, especially for autonomous manipulation.	✗	✗	✓	✓	V, L, O
	2026	*Ours*	2020–2025	Systematic review of core EAI approaches for industrial applications, providing a holistic overview of physical embodiments, input/output modalities, datasets, and learning mechanisms to guide practical and academic development.	✓	✓	✓	✓	✓

Summary of the thematic scope of recent surveys and reviews. The column 'Industrial Application' indicates whether industrial use cases are discussed; P denotes partial coverage. 'Embodiment' indicates whether different physical embodiments are explicitly considered or compared. 'Learning' categories surveyed by the review papers in the context of EAI are RL — reinforcement learning, IL — imitation learning, F — foundation models such as LLMs, VLMs, or VLAs, and O — other learning approaches. 'Input Modalities' are V — vision, L — language, T — tactile, O — other, and MM — multimodal input. Listed abbreviations indicate categories explicitly addressed by the review; omitted categories are outside its scope or not analyzed separately. 'Datasets' indicates whether datasets are explicitly addressed. A ✓ in the 'Learning' or 'Input Modalities' column indicates that the paper covers all listed categories for that column.

2. Fundamentals

Before presenting the SLR, we establish a common understanding of what constitutes EAI in the scope of this work. We first present the historical evolution of the term EAI, outlining its definition for the remainder of this work. We then examine the architecture of an EAI system. Further, we explore the diverse spectrum of physical morphologies an EAI system can inhabit by classifying them along a manipulation-mobility matrix. Together, this section provides the necessary terminology and structure for the subsequent SLR to evaluate current EAI approaches in industrial operations.

2.1. Terminology and characteristics

In recent years, EAI has emerged as a concept across diverse research domains, ranging from cognitive science to foundation models and advanced robotics. However, this widespread adoption has led to a significant fragmentation of its definition. Different authors frequently employ the term to describe entirely different paradigms, often deviating substantially from its original intended meaning. For instance, some modern interpretations use EAI merely as a buzzword to denote any AI deployed on hardware, neglecting the foundational principle that intelligence should fundamentally emerge from the interaction between a physical body and its environment. To navigate this ambiguity and establish a baseline for this SLR, we examine the historical evolution of the concept. By tracing EAI back to its theoretical origins, we ground the definition for the rest of this work.

Already in 1950, Turing argued that, rather than trying to program an adult-level intelligent machine, one could produce a machine that simulates a child's mind and educate it until it reaches adult-level intelligence. During education, rewards and punishments shape the child's behavior. However, the first attempts to build and teach a child-level program will likely fail, which is why multiple attempts are required, and the outcomes must be judged. This closely follows how evolution works. He further concludes that instead of structurally searching for the best solution, introducing randomness inside the machine imitates mutations, which, faster and less computationally expensive, leads to a more satisfactory solution, also known as evolutionary search [20]. Turing describes the basic building blocks of modern EAI without specifically using the terms. Later, in 1987, R. Brooks states that human-level intelligence cannot be decomposed into modelable subpieces, including also that we have no deeper knowledge of the interfaces between them [21]. Instead, he concludes that intelligence can emerge from incrementally learning by being and reacting. L. Smith formulates that into the embodiment hypothesis, stating that "intelligence emerges in the interaction of an agent with an environment and as a result of sensorimotor activity". She relates this to how a baby learns to reach human-level intelligence at one point [22], which closely follows Turing's idea.

While Brooks uses the noun embodiment, Smith speaks of embodied intelligent agents; the definition of the specific term embodied (artificial) intelligence is ambiguous, or, in other words, the definition of what a system makes embodied is described differently [23]. Table 2 lists various EAI "definitions" and characteristics. In early works, the term EAI was used epistemologically and formed methodologies of how to build an intelligent system, including a heavier influence of biologically inspired and materially grounded approaches [24–26]. Later, starting with the deep learning revolution in the mid 2010s, the term was especially used to describe the physical presence compared to virtual intelligent systems, while, after the rise of foundation models, using it as a distinction from "Internet AI", where learning solely happens on virtual data, gained importance [27].

Regardless of early or later works, an EAI system has some form of embodiment situated in the real world. Further, intelligence emerges from interaction with the environment as described in both early [22, 25] and most recent [13] works. This characteristic differentiates EAI from Physical AI, where the coupling of the embodiment and learning is not mandatory. Thus, Physical AI is an overarching category, defining an intelligent system that has some physical representation, regardless of an embodiment, which allows for a higher level of intelligence through learning. As an example, a vision system that detects specific objects in camera images, which is mounted on a robot, will not reach a higher level of intelligence when not allowed to learn in the specific embodiment, e.g., learning that moving the camera to the left or right will allow for a larger field of view.

While some of the recent works define EAI with a single describing broader sentence [15,23,28], others allow fulfilling multiple characteristics to different extents when defining an EAI system [13,29], as detailed in Table 2. We will follow the latter approach and define EAI in terms of specific characteristics that describe what makes a system embodied intelligent. Thus, we can review the literature with its slightly different EAI definitions and characteristics, and map them to ours. The criteria will therefore be used to systematically review the literature. Rather than checking for "hard" requirements, we will assess the degree to which each paper addresses the specific dimensions of EAI. We intentionally depart from definitions that equate embodied intelligence with biologically inspired mechanisms, or other integrations, e.g., cognition is integrated within physical entities, as some works discuss [24,25], since these approaches and concepts are far apart from ongoing research and development around real-world use cases in the industrial domain. Based on the previous outline, historical developments, and manifestations of the term EAI, we will use the following four characteristics for the rest of this work:

- **(A) Embodiment**: The necessity of a physical or simulated body. Without a body, intelligence remains "disembodied" (e.g., pure LLMs).
- **(B) Learning**: Besides the embodiment, learning is the fundamental module required to achieve a higher level of intelligence, as already described by Turing [20]. If there is no learning within an environment or within the embodiment, the field transitions into Physical AI, e.g., a trained perception stack controls a robot, while the type of robot is exchangeable. We further distinguish between:
  - **(B.a) Embodied Learning**: Focuses on learning within the constraints and capabilities of the specific morphology (e.g., motor skills, joint limits). This often requires high-fidelity simulation or real-world training.
  - **(B.b) Environmental Interaction**: Learning through active engagement with external stimuli. The distinction between (a) and (b) allows for a more granular analysis of how an agent acquires knowledge.
- **(C) Perception and Action Coupling**, which is the result of Brooks's [21] and Smith's [22] EAI hypothesis, as outlined above.
  - **(C.a) Dynamic Sensorimotor Feedback**: Perception is not static; it changes as a direct result of the agent's actions (e.g., an "ego-centric" camera mounted on the robot vs. a static ceiling camera). This includes tactile sensing and proprioception.
  - **(C.b) Model-Level Integration**: The coupling occurs within the architecture itself. The "intelligence" between perception (input) and action (output) is deeply integrated and not a (set of) modular, swappable "planner" blocks as seen in classical robotics.
- **(D) Situated Intelligence**: The ability of an agent to understand its environment in a specific context [13]. Example: A vacuum robot demonstrating situatedness does not merely perceive a "gap" (doorway); it understands the semantic transition of moving from one room into another one. This implies a higher level of spatio-temporal awareness.

Table 2
Uses and meanings of EAI in literature.

Authors	Year	Title	EAI definition
R. Pfeifer et al.	2003	Embodied Artificial Intelligence: Trends and Challenges [24]	EAI is defined as the paradigm using the three-step method: (1) understanding biological systems, (2) based on it, abstracting general principles of intelligent behavior, and (3) applying the gathered knowledge to build an intelligent system.
R. Chrisley	2003	Embodied artificial intelligence [25]	Outlines that there are different approaches to EAI based on the notion of the embodiment: sole physical realization and physical/organismoid/organismal embodiment, which classification stems from cognitive science [30].
T. Froese et al.	2009	Enactive artificial intelligence: Investigating the systemic organization of life and mind [26]	EAI is presented as a paradigm that moves away from traditional, computationalist AI and instead focuses on agents that are physically situated in the real world. The authors present multiple design principles that EAI practitioners use. These are more broadly, including, e.g., "parallel, asynchronous, partly autonomous processes, [which are] largely coupled through interaction with the environment". Further, they discuss enactive AI, which is named as an evolution of EAI. In enactive AI, rather than adding a value system to the AI that "motivates" it, an enactive AI system has an intrinsic motivation to learn.
N. Roy et al.	2021	From Machine Learning to Robotics: Challenges and Opportunities for Embodied Intelligence [23]	Takes the definition from [31] for robotics and alters it to define EAI as "the purposeful exchange of energy and information with a physical environment".
Y. Ma et al.	2025	A Survey on Vision-Language-Action Models for Embodied AI [28]	Describes EAI as "a unique form of artificial intelligence that actively interacts with the physical environment", whereas robots are one of the most prominent embodiments within the field.
Y. Liu et al.	2025	Aligning Cyber Space With Physical World: A Comprehensive Survey on Embodied AI [15]	Defines EAI by its operation within physical space, where cognition is integrated into physical entities such as robots, cars, and other devices, compared to disembodied AI, where cognition and physical entities are disentangled.
L. Ren et al.	2025	Embodied Intelligence Toward Future Smart Manufacturing in the Era of AI Foundation Model [29]	Embodied intelligence exhibits the characteristics: continuous learning and evolution, open-world generalization, physical multimodal interaction, collaborative swarm intelligence, and simultaneous signal processing.
H. Liu et al.	2025	Embodied Intelligence: A Synergy of Morphology, Action, Perception and Learning [13]	EAI is defined as the "synergy of morphology, action, perception, and learning". Among these, the authors present eight core principles, demonstrating the bidirectional influence of the components, such as how an agent's physical form can offload computational tasks to simplify action control, or how active exploration enhances environmental perception.

We adopt these four characteristics as the working definition of EAI for this review, treating them as graded dimensions rather than binary requirements: a system can be embodied intelligent to varying degrees, and each reviewed study is assessed by the extent to which it realizes them.

2.2. System components

An EAI system forms a multi-layered architecture designed to translate, possibly, human objectives to physical execution. As illustrated in Fig. 2, we decompose the architecture into context boundaries, hardware components/interfaces, modular software layers, and enabling components. To illustrate these components practically, we adopt a common household task as an example throughout this section: a user verbally instructing a domestic mobile manipulator to *"Put the coffee mug in the sink"*.

The **context components** establish the operational scope and boundary conditions for the system. Their primary goal is to formalize the problem space that the EAI must navigate by taking the user's raw intent and the physical constraints of the real world as inputs and mapping them to actionable outputs. In the example given in Fig. 2, the task instruction is the verbal command to move the mug, while the work environment is the unstructured kitchen containing various obstacles and lighting conditions. In the literature, these environments are often formalized within standardized evaluation benchmarks such as ALFRED [32] or BEHAVIOR-1K [33].

Hardware components act as the physical boundary mediating information transfer between the environment, the user, and the software components. The user and perception interfaces aim to capture external human intention and actions as well as physical phenomena. For instance, a microphone records the user's voice command, while an RGB-D camera captures point clouds of the kitchen table.

The **software** stack serves as the cognitive pipeline, processing the ingested data to compute physical agency. Depending on the system design, these layers can be strictly modular, partially coupled, or fully end-to-end integrated, as seen in models like RT-2 [34], OpenVLA [35], and π_0 [36].

First, the interaction and perception layers translate the raw sensor and text inputs into, e.g., structured semantic task representations or other embeddings. Next, the cognition and planning layer combines the task objective with the environmental state by decomposing the high-level task into a sequence of actionable primitives or states. Finally, the action layer computes the continuous low-level control trajectories required to transition between sub-states.

In the scope of manufacturing, the outlined decomposition may show similarities to other frameworks like Cognitive Digital Twins (CDT). A CDT is described as "a digital representation, augmentation, and intelligent companion of its physical twin as a whole", which is "augmented with certain cognitive capabilities and supports executing autonomous activities" [37–39]. However, Digital Twins (DT) as digital replicas of real-world, physical assets, and the cognition-labeled extension, are conceived in a fundamentally different approach. A CDT is constituted by a twinning relation: it is the digital counterpart of a specific physical asset, process, or system, and its cognition is grounded in models, ontologies, and process knowledge of that referent. The digital part, the boundaries of cognition, reasoning are modeled by a human. An EAI system is thought without modeling a virtual replica; it is the embodied agent itself, and by our working definition (Section 2.1), its intelligence must be acquired through learning that is constrained by its morphology and driven by interaction with its environment (criteria B and C). By definition, the digital counterparts of DTs and CDTs are crafted by how humans would decompose systems and intelligence into individual blocks, whereas the EAI hypothesis demands that the system learn this by itself, e.g., being just one large Neural Network (NN) without any modeled priors on how the world works and interacts, which is instead learned during training.

Further **enabling components** provide infrastructure and data necessary for training prior to real-world deployment. Generalist policies like RT-1 [40] and OCTO [41] are trained on massive multi-robot datasets like Open X-Embodiment [42], while physics-accurate simulators (e.g., NVIDIA Isaac Lab [43], SAPIEN [44], Meta Habitat [45–47]) facilitate safe Reinforcement Learning (RL) and sim-to-real transfer. Beyond large-scale dataset curation, other approaches lower the data burden by transferring human skills directly to the robot, for instance

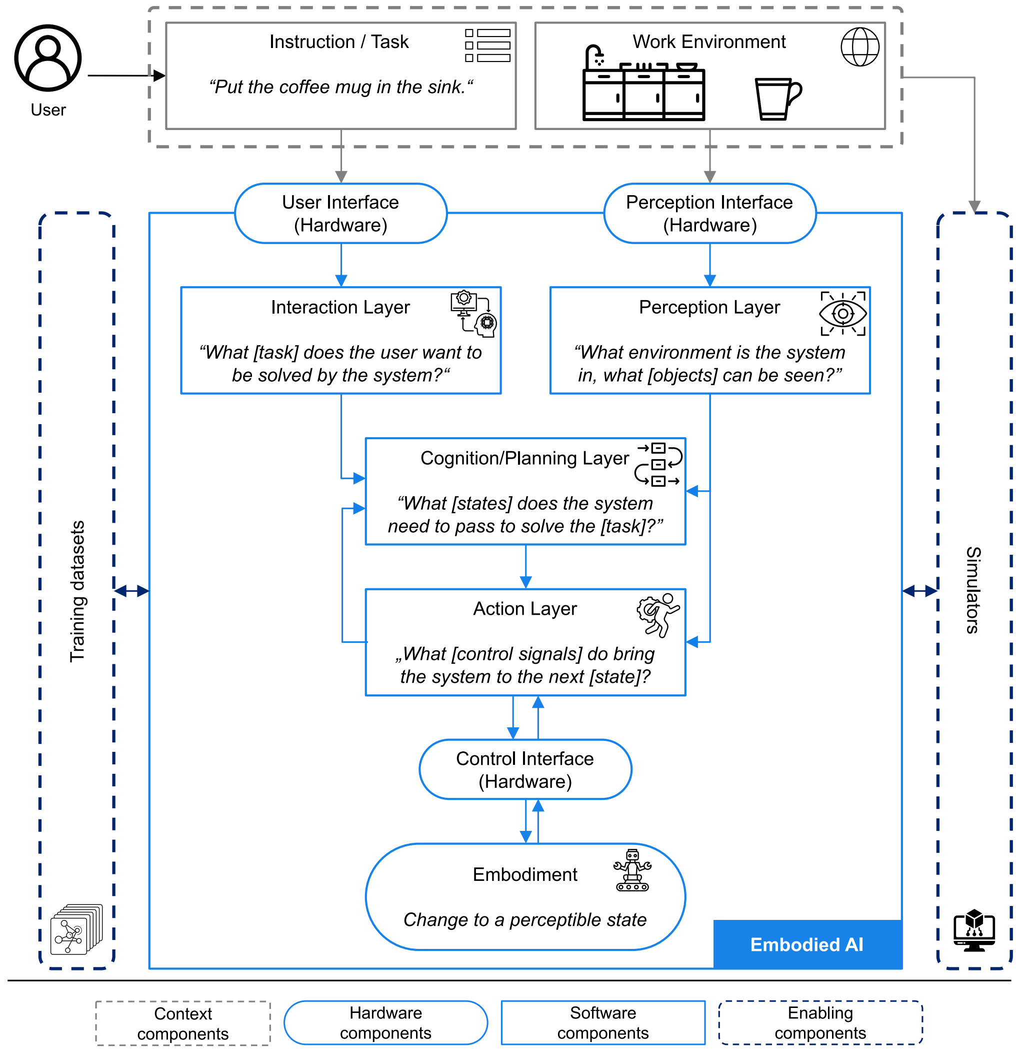

Fig. 2. EAI building blocks.

via mixed-reality-assisted demonstration of visuomotor primitives for contact-rich assembly [48].

Once the cognitive pipeline has generated the necessary control signals, the information flows back into the physical world through the executing hardware. The control interface and embodiment are responsible for realizing the computed physical changes. The control interface translates the digital commands output by the action layer into physical actuator state changes.

2.3. Physical embodiments

The primary defining characteristic of EAI is the embodiment (see Section 2.1), distinguishing it from disembodied AI. While public discourse often equates EAI with humanoid robots, the reality encompasses a much wider variety of systems.

To categorize these diverse physical forms, Fig. 3 maps EAI systems across two primary dimensions: mobility (ranging from stationary to unstructured terrain navigation) and manipulation (ranging from zero physical interaction to highly dexterous object handling). Based on the numbered clusters in Fig. 3, these embodiments are as follows:

- **Cluster 0 (No Embodiment):** Purely digital entities like avatars or chatbots. Included here to explicitly mark the boundary of physical embodiments.
- **Cluster 1:** Automated Guided Vehicles (AGVs) and social robots. These navigate flat, structured environments, e.g., to transport goods (requiring external loading), or to interact socially without altering their physical surrounding.
- **Cluster 2:** Unmanned Aerial Vehicles (UAVs) and quadruped robots without grippers. Such systems are highly mobile, and can be deployed for tasks such as visual inspection across complex terrains or in the air.
- **Cluster 3:** Traditional industrial robotic arms with simple grippers. These systems are stationary and can handle, e.g., standard manufacturing tasks like palletizing, logistics commissioning, and machine tending.
- **Cluster 4:** Mobile robots and automated forklifts. Equipped with simple grippers or levers, these systems typically dominate intralogistics by retrieving and transporting standard objects across factory floors.
- **Cluster 5:** Humanoids, UAVs, and quadrupeds with basic grippers. Designed for uneven terrain or outdoor tasks, such as construction site logistics or delivery.
- **Cluster 6:** Stationary industrial arms with complex, multi-fingered end-effectors, or multi-robot setups. These can, e.g., handle highly dexterous manipulation tasks involving complex, deformable, or oversized objects.
- **Cluster 7:** Mobile manipulators with advanced grippers. These systems can perform dexterous manipulation tasks across multiple static workstations, optimizing indoor operational efficiency.
- **Cluster 8:** Highly dexterous humanoid robots. Representing the upper limit of current mechanical complexity, these can navigate

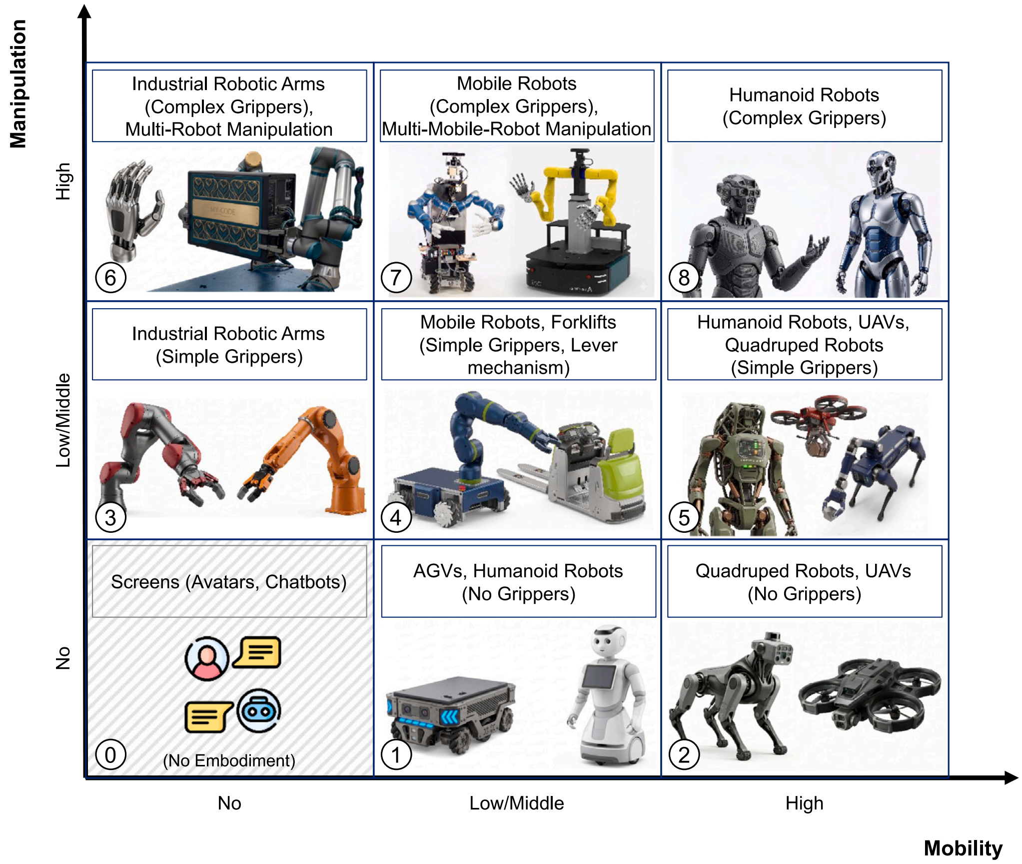

Fig. 3. Manipulation-Mobility Matrix clustering different physical embodiments and their skills (robot images were created with generative AI).

unstructured environments (such as climbing stairs in households) while performing fine-motor skills.

3. Review method

The SLR is based upon the guidelines and recommendations from [49,50]. We document it in Fig. 4 using a PRISMA flow-chart [51]. Before conducting the review, we began with initial planning and defined the scope using the PICO framework to formulate research questions (RQs). This was followed by the search string composition used to query relevant records. We defined and applied study selection criteria to reduce the body of literature while moving further down in the PRISMA flow-chart. Before the data extraction, we additionally assessed the paper's relevance to our study using a set of assessment questions.

3.1. Research scope using PICO

The PICO (Population, Intervention, Comparison, Outcome) framework, originally used in evidence-based medicine research, can be used to break down the problem at hand and help with formulating research questions to allow a comprehensive and focused search strategy. We use the list of PICO elements and their corresponding terms in computer science compiled by [49] and identified the relevant keywords and synonyms for each PICO component relevant to our study in Table 3. Please note that for the **I**ntervention component, we also added the term "Humanoid" as a synonym for "Embodied AI", as it is likely the most common embodiment in the field.

3.2. Research questions

Based on the PICO scope, the following research questions (RQs) were formulated to guide this review. The phrases printed in **bold** are those taken from the PICO keyword collection in Table 3:

1. How is the **term EAI** characterized, defined, and applied within the current scientific literature on industrial systems?
2. What are the primary **research focuses, use cases, and industrial applications** of EAI?
3. What **physical embodiments** are primarily utilized for EAI systems in industrial contexts?
4. How are the core **learning components** coupled within system architectures, and what **input and output modalities** are used/implemented?
5. To what extent are **simulation environments** utilized, and what are the prevalent **training strategies**?

3.3. Search strategy

3.3.1. Relevant databases

To ensure comprehensive coverage, we conducted the search across the following academic databases:

- Scopus
- IEEE Xplore
- arXiv (for pre-print articles)

3.3.2. Search string composition

The final search string used to obtain the initial corpus of literature in PRISMA workflow step 1 was composed using the keywords identified in the PICO analysis. We selected the most representative

Table 3
PICO keyword collection for search string composition.

Component	Description	PICO keywords	PICO keyword synonyms
Population/**P**roblem	Application area, industry domain selected for this study.	Industrial Manufacturing/Industrial Operations.	Manufacturing; Assembly; Maintenance, Repair, Overhaul; Inspection; Production/Manufacturing Systems; Logistics; Industrial Environments; Industry 4.0; Industry 5.0
Intervention	Technology being investigated.	Embodied AI.	Embodied Intelligence; Agentic AI; AgentAI; Intelligent Robotics; Physical AI; Autonomous Systems; Artificial General Intelligence (AGI); Embodied AGI; World Models; Vision-Language-Action (VLA) Model; Humanoids
Comparison	Currently applied approaches (alternatives to the Intervention).	Traditional automation and robotics, manual processes.	Automation; Fixed Robotics; Manual Assembly; Non-collaborative Robots; Industry 3.0; Automated Assembly; Automated Manufacturing; Automated Logistics
Outcome	The results and effects the studies produce or reveal.	Applications, benefits, evaluation metrics, and potentials.[a]	Task; Use Case; Implementation; Benefits; Challenges; Limitations; Future Trends; Performance Indicators; Productivity; Efficiency; Flexibility; Safety; Ergonomics

[a] We are investigating the current state of the art in EAI, its applications, challenges, and future directions. Therefore, the **Outcome** is not treated as a single keyword here, but rather a collection of intended outcomes we aim to provide through this study.

keywords and synonyms from Table 3 for both PICO components **P**opulation/**P**roblem, and **I**ntervention to combine them in their respective groups using a Boolean OR operator. We then linked the two groups using a Boolean AND operator to compose the final search string. The selected terms are marked **bold**:

- Population/Problem: Industrial Manufacturing/Industrial Operations, Manufacturing; **Assembly**; **Maintenance, Repair, Overhaul**; Inspection; **Production**/Manufacturing **Systems**; **Logistics**; Industrial Environments, Industry 4.0, Industry 5.0
- Intervention: **Embodied AI**; **Embodied Intelligence**; **Agentic AI**; AgentAI, **Intelligent Robotics**; **Physical AI**; Autonomous Systems; **General Artificial Intelligence** (AGI); **Embodied AGI**; **World Models**; **Vision-Language-Action** (VLA) Model; **Humanoids**

Composed Search String:

```
("Manufacturing" OR "Assembly" OR "Logistic*" OR "Production
System*" OR "MRO" OR "Maintenance, Repair and Overhaul" OR
"Maintenance, Repair, Overhaul") AND ("Embodied AI" OR "Embodied
Intelligence" OR "Agentic AI" OR "Intelligent Robotics" OR
"Physical AI" OR "General Artificial Intelligence" OR "Embodied
AGI" OR "World Models" OR "Vision?Language?Action" OR "Humanoid*")
```

As each search engine requires a slightly different composition of the search string, searchable fields, and other parameters are given in Appendix A.1. Because quoted search terms occasionally returned results containing only individual words (e.g., "world" and "models" instead of "world models"), we applied a post-search filter step prior to inclusion in the PRISMA workflow, see Appendix A.1. After retrieval and filtering, 554 SCOPUS records, 230 IEEE Xplore records, and 79 arXiv records remained as input to PRISMA workflow step 2.

3.4. Study selection criteria

We defined selection criteria for the literature to be included to further refine the body of literature initially queried. The criteria for each category and the corresponding PRISMA step in which we applied the criteria are compiled in Table 4. The timeframe constraint (PRISMA step 1) was directly applied when gathering records from the relevant databases described previously in Section 3.3. To further reduce the amount of literature in 2, we have decided to exclude literature that did not make a sufficient impact. To achieve this, we applied a median-based citation count filter that rejects records that have a citation count less than the median citation count of all papers published in the same year, evaluated individually for each source to avoid inequalities due to different citation count retrieval strategies used by the three providers. For publications dated 2025 and 2026, this criterion was not applied to avoid rejecting huge amounts of new literature that has not been cited yet due to its novelty. In 3 and 4, we manually checked the literature type, availability, and language before conducting a more detailed relevance screening in 5 and 6. In 5 we checked title, abstract, and keywords to reject irrelevant papers, taking into consideration the RQs we are addressing, see Section 3.2. Followingly, in 6 we applied a relevance assessment checklist, further detailed in the following Section 3.5.

3.5. Relevance assessment checklist

In [50], a Quality Assessment (QA) checklist is proposed as the last screening step before data extraction to evaluate the quality with regard to the scope of the survey. Given the heterogeneity of research in EAI, standard quality checklists are not directly applicable, as we are trying to provide a broad overview in the field of EAI related to industrial manufacturing, despite criteria such as general relevance or methodological design. At this point, we introduced four Relevance Assessment (RA) questions that we used in PRISMA step 6 to evaluate the relevance to our study topic. Each criterion was assessed by all authors using a binary scoring system. Any discrepancies in scoring were resolved through discussion to reach a consensus. This process ensured a transparent and repeatable assessment of the included literature's quality. We considered a paper for inclusion if all RA questions could be answered with "yes".

RA1: Industrial context: Is the paper linked to and relevant for an industrial context? Does the study explicitly address or demonstrate relevance to an industrial setting (e.g., manufacturing, logistics, MRO), or is its applicability clearly articulated or discussed?

RA2: Machine Learning: Does the paper apply Machine Learning (ML) as a key component of their proposal?[2]

RA3: Existence of some form of embodiment: Does the contribution revolve around an embodiment being present either in the physical world or in a simulated environment? We aim to include only records that explicitly deal with some form of embodiment and reject contributions that are, e.g., vision- or language-only without any action layer, see the EAI definition criteria (A) and (B) in Section 2.1.

RA4: Character of the paper: We aim to investigate the use, technologies, challenges, and future research directions for EAI in an industrial context. We therefore examine the character of

[2] To narrow down the term AI, we focus on data-driven EAI concepts that employ machine learning. This delimitation ensures conceptual clarity and allows us to focus on approaches characterized by learning from data.

Table 4
Study selection criteria for inclusion/exclusion.

Category	Criteria	PRISMA step
Timeframe	**Exclusion:** Studies published prior to 2020 or not within the queried databases on December 5th, 2025 (time of freezing the search).	1
Impact	**Exclusion:** Consideration of literature whose citation count is at least equal to the median citation count of all queried papers in the corresponding year, evaluated individually for each database. This criterion is not applied to publications dated 2025 and 2026.	2
Literature Type	**Inclusion:** Peer-reviewed and pre-print journal and conference articles Exclusion: Editorials, opinion pieces, review papers, and non-technical articles.	3 + 4
Availability	**Exclusion:** Full-text not available.	4
Language	**Inclusion:** English language publications only.	4
Relevance	**Inclusion:** Literature that is linked to and discusses EAI within the intended scope of this review. In a first relevance assessment using only title, abstract, and keywords in 5 we use the research questions formulated in Section 3.2 for guidance. In a second full-text relevance assessment in 6 we apply the four RAs further detailed in Section 3.5.	5 + 6

the paper and evaluate if it contains some form of implementation, practical investigations with evaluations, etc. We excluded contributions dealing merely with, e.g., completely conceptual proposals or philosophical discussions.

3.6. Data extraction and synthesis

The number of records assessed following the PRISMA workflow is given in Fig. 4. A list of the 166 records was assessed for eligibility through a full-text screening in the last PRISMA workflow step 6. In this step, we excluded 15 records due to multiple reasons (review, non-English, duplicate, or no fulltext available) that did not become obvious in the previous steps, in which fulltext was not assessed. The resulting list of 151 records assessed using the RA checklist is presented in Table 5. During this step, another 109 records were excluded due to irrelevance, as at least one RA question was answered with "no". This leaves us with a total of 42 records that are included in this review. Each included record is also marked in Table 5.

4. Results

This section presents the findings of our SLR. We first provide an overview of the general research landscape by assessing the usage of the term EAI against our proposed definition, alongside mapping the key research focuses, industrial applications, and physical embodiments. Following this landscape analysis, we dive deeper into the underlying technological approaches. By evaluating the coupling of learning components, multimodal integration, the role of simulation environments, and prevailing training paradigms, we provide an overview of the current deployment of EAI systems for industrial environments.

4.1. Use of the term EAI

Across the reviewed papers, the term and characteristics of EAI in the industrial domain are used inconsistently. This is consistent with other literature, as already outlined in Section 2.1.

Many works adopt the EAI label, while always fulfilling the criteria of having a physical or simulated embodiment, as already checked ealier using the relevance assessment question RA3 (see Section 3.5), multiple works lack learning mechanisms within the embodiment and environment (criterion B, as defined in Section 2.1), e.g., [52–54].

Further, no work elaborates on the different EAI definitions and characteristics. Beyond these terminological discrepancies, we reveal the following three trends.

Morphological constraints driving EAI (criterion B.a). When works use non-standard morphologies, learning in the embodiment is more frequently used than when works use standard morphologies, such as industrial robots or even humanoids. Works featuring highly specific physical systems, such as thruster-aided humanoids requiring specialized controllers [55], often involve unconventional embodiments that force a significantly tighter coupling between the agent's morphology and its learning process, thereby adhering more closely to the strict definition of morphological computation.

However, this conclusion is biased by the fact that for standard morphologies, a larger number of existing models are available that can be used to generate actions. When choosing a non-standard morphology and deciding to employ, e.g., RL, it is evident that embodied learning is fulfilled.

Foundation model paradox (criterion C.b). Recent works increasingly include LLMs and VLMs to drive perception and reasoning [52,56]. However, when integrating these and using human-understandable inputs and outputs, these building blocks become swappable and contradict the model-level integration hypothesis (C.b). Many works use LLMs as high-level, modular planners that generate, e.g., text-expressed robot primitives [52,54,57]. This modularity, in which the cognitive reasoning is detached from the specific physical constraints of the embodiment, deviates from the integrated, non-swappable sensorimotor coupling as underlined in many EAI definitions.

Missing situated intelligence (criterion D). While a subset of the reviewed literature successfully demonstrates a complete perception-action loop (C) coupled with embodied or environmental learning (B) [58,59], there is a notable absence of situated intelligence (D). The majority of systems translate sensor inputs into motor outputs, e.g., extracting targets from egocentric cameras to generate actions (C.a) [60,61]. However, these systems largely operate as reactive executors, devoid of semantic and spatio-temporal awareness. The systems interact with their environment mechanically, but most works fail to demonstrate a broader contextual understanding of their actions.

While the mechanical and learning foundations of EAI are increasingly entering the industrial domain, much of the existing work shows inconsistent engagement with its definitions, meanings, and defining characteristics.

On the one hand, this is necessary to establish common terminology; on the other hand, it enables researchers and practitioners to benefit from the ideas behind EAI when implementing such systems. Otherwise, approaches remain within broad conceptual labels rather than taking a step further from, e.g., Physical AI to EAI.

4.2. Key research focuses

We categorize the surveyed literature into distinct thematic clusters. This gives an overview of the literature's primary focus areas. Specifically, we partition the current research into six core clusters, see also Fig. 5:

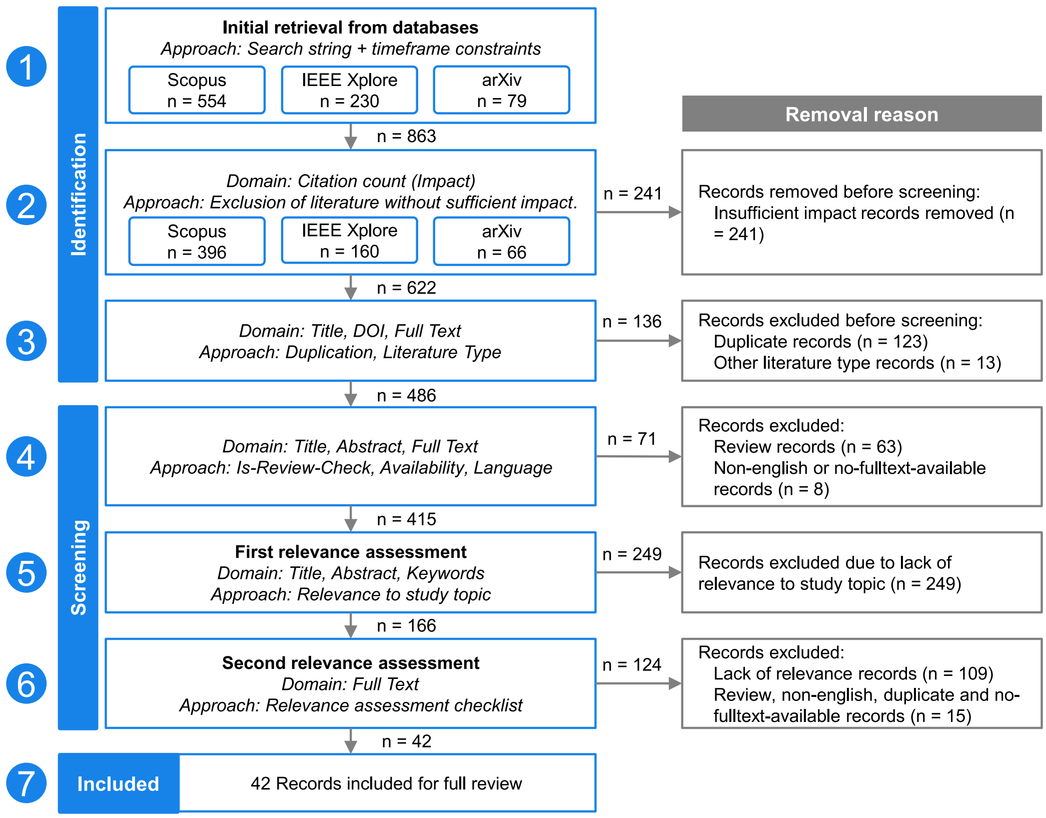

Fig. 4. PRISMA flowchart. The search string, other constraints and criteria as well as the relevance assessment questions are further detailed in Sections 3.3–3.5 and Table 4.

- **Cluster 1: Human-Robot Collaboration (HRC):** Transforming robots into proactive partners that can work in collaboration with humans, utilizing, e.g., multimodal human information.
- **Cluster 2: Foundation Model-Based Decision Making, Planning, Reasoning, and Control:** Creating frameworks that map high-level instructions and visual inputs to executable robot control signals, and long-horizon tasks, or approaches that infer how to reach a given goal.
- **Cluster 3: Autonomous Manufacturing, Machining, and Assembly Tasks:** Automating specialized, domain-specific industrial processes.
- **Cluster 4: Robotic Grasping, Dexterity, and Manipulation:** Papers dealing with the physical mechanics of interaction, focusing on dexterous manipulation and collision-free grasping.
- **Cluster 5: Mobility, Navigation, and Multi-Agent Systems:** Enabling autonomous agents to navigate and self-organize themselves in dynamic industrial environments.
- **Cluster 6: Datasets and Data Generation:** Research that presents domain-specific datasets, proposes data-collection protocols, and addresses sim-to-real challenges.

Cluster 1: Human-Robot Collaboration (HRC). In the first cluster, the authors approach the question of how to transform robots from passive participants into active collaborators and deal with topics revolving around how to improve joint collaboration and mutual perception between humans and robots. For instance, [62] discusses how to enable compliant human-humanoid collaboration for collaborative object carrying to reduce the physical strain of the human while relying solely on proprioceptive information. Their approach includes the usage of a residual teacher policy and distillation. Another example in this cluster is [63], in which the authors design a proactive collaborative controller to help a robot transition from a passive "follower" role to a "collaborator" role in unstructured, unknown scenarios. The controller uses human multimodal information (joint angles and surface electromyography (sEMG) signals) to predict human arm motion and learn variable impedance behavior for stable and comfortable collaboration. The question of how HRC assembly can be made more adaptive, intelligent, and safe while operating in complex, evolving environments is discussed in [57,64]. A digital twin simulation helps to verify the generated code before execution.

Cluster 2: Foundation model-based decision making, planning, reasoning, and control. The second cluster covers various foundation models (VLAs, LLMs) and topics revolving around decision making, planning, reasoning, and robotic control. For example, [56] discusses how VLMs can be employed for assembly sequence generation from technical documents and assembly drawings to achieve zero-shot assembly sequence learning through a multi-hierarchical scene graph architecture that allows open-set product assembly. In [65], the authors investigate how to enable VLAs to conduct long-horizon tasks and employ a unified framework using a Mixture-of-Transformers (MoT) architecture and Chain-of-Thought (CoT) reasoning to generate multimodal "manuals" for robot execution that decompose the task into smaller portions. Long-horizon tasks also motivate [58], where such tasks, often characterized by sparse-reward environments, are tackled by the introduced pipeline

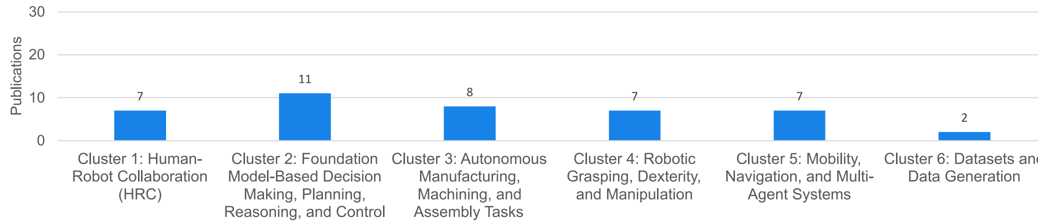

Fig. 5. Surveyed literature's six focus research clusters.

that employs pre-trained LLMs to generate dense reward functions and thus improve subsequent learning performance. [66] investigates how robots can acquire dexterous skills from human demonstration. [67] focuses on generating high-precision 3D physical grounding without model retraining through the coupling of multiple models (VLMs and LLMs). The VLM performs general reasoning and task planning, while the LLM ensures the motion is logical. [68] deals with the question of whether optimizing visual and text prompts enhances their spatial reasoning enough to automate robot failure recovery.

Cluster 3: Autonomous manufacturing, machining, and assembly tasks. The third cluster contains literature revolving around autonomous manufacturing, machining, and assembly tasks. As planning and decision-making form a major part in enabling such tasks, the approaches slightly overlap with the second cluster. In [69], the peg-in-hole challenge is addressed using a procedural generation approach to train an RL agent. Peg-in-hole tasks are common across various assembly operations, and the authors tackle this problem by leveraging synthetic environment generation to train an RL agent across diverse environments. The potential of LLMs for industrial design, decision-making, and tool-path generation is addressed in [52]. The multi-stage framework matches a task description with process parameters and generates corresponding paths, which are then executed in a simulation environment. Their evaluations showcase various tasks such as gripping, placing, and stacking objects. [70] proposes a behavioral planning and parameter learning approach that utilizes a low-code/no-code paradigm and decomposes the given task using a multi-stage framework.

Cluster 4: Robotic grasping, dexterity, and manipulation. The fourth cluster deals with robotic grasping, dexterity, and manipulation. In [71], the placement of multiple irregularly shaped objects in an arrangement that is as compact as possible is investigated. The authors propose an adjacency-aware RL approach that utilizes reference lines to maintain boundary awareness and apply a gap-minimizing reward function. The question of how robot arm grasping performance can be improved through enhanced world models is addressed in [72]. The authors introduce a backward-dynamics prediction module to capture temporal dynamics and improve sample efficiency. The design of a humanoid hand and the presentation of a tailored data collection protocol are presented in [73]. The paper discusses how robust data collection and a curation protocol can enable policies to recover from failures and generalize across diverse dexterous manipulation tasks. Their strategy includes heavy randomization and strategic inclusion of self-correction data that is tested with the designed humanoid hand for grasping and manipulation. A new dataset and background-adaptive grasping network are introduced in [74]. The authors discuss how robot grasp detection can be improved and present a new RGB-D cross-background dataset to enable grasping in cluttered scenes. Tightly coupled with hardware control, [75] investigates how to control a humanoid robotic hand actuated by shape memory alloy with a model-free strategy.

Cluster 5: Mobility, navigation, and multi-agent systems. The fifth cluster focuses on mobility, navigation, and multi-agent systems. In [76], the authors investigate how navigation accuracy, obstacle avoidance, and manipulation can be improved for indoor mobile robots. They combine multiple sensing modalities to facilitate situation awareness and intelligent task scheduling. The question of how to achieve self-organization in heterogeneous multi-agent systems is discussed in [77]. The proposed system integrates multimodal data with VLMs and LLMs for planning and orchestration. Similarly, in [78], the coordination mechanism and a learning strategy for fully cooperative multi-agent systems in humanoid robotics are investigated. In [59], the question of how to make humanoids move more like humans, including typical bottlenecks such as motion discretization, strategy homogeneity, and response delays, is addressed.

Cluster 6: Datasets and data generation. The last cluster deals with datasets and data generation. In [79], the authors investigate how to segment robotic actions into primitives for better VLA training. Their intent is to support preparation for VLA-pretraining from larger amounts of data. Synthetic training data generation for developing generalist policies for robot manipulation using high-fidelity synthetic data is presented in [80]. The authors present a pipeline to generate synthetic training data such as motions, renderings, and instructions. As training data is the basis for enabling other tasks grouped into other clusters, there is also an overlap among them.

This section systematically categorizes the surveyed literature into six primary research clusters to highlight current focus areas, providing a thematic overview of all papers that satisfy the relevance assessment criteria (see Section 3.5). Interestingly, the distribution reveals that works around HRC are represented equally in numbers to both autonomous manufacturing, machining, and assembly, as well as robotic grasping, dexterity, and manipulation. This indicates a balanced focus across the field: foundational research (such as grasping and manipulation) appears in equal measure to both human-centric applications and highly specific industrial operations. However, works centered on the integration of foundation models occur slightly more frequently than the others.

4.3. Industrial applications

Across the reviewed literature, industrial application areas (Fig. 6) are distributed unevenly: assembly is the most prominent application area, followed by manufacturing- and process-oriented operations, while logistics appears regularly but less frequently, and inspection and maintenance-related scenarios remain comparatively rare (see Fig. 7). A noticeable share of papers also cannot be assigned to a single clearly bounded industrial domain, as they focus on generic frameworks, datasets, or architectures with only loosely specified industrial relevance.

Assembly. Assembly is frequently addressed as a key motivation for industrial EAI because it combines industrial relevance with clearly operationalizable subproblems such as grasping, insertion, sequencing, and recovery [65,69,81]. A major theme is high-level task planning and long-horizon execution, where VLMs or LLMs decompose assembly goals into executable subtasks and connect perception with decision making across extended action sequences [56,58,65]. Assembly is therefore relevant not only because it involves physical manipulation, but also because it requires structured reasoning over multi-step processes.

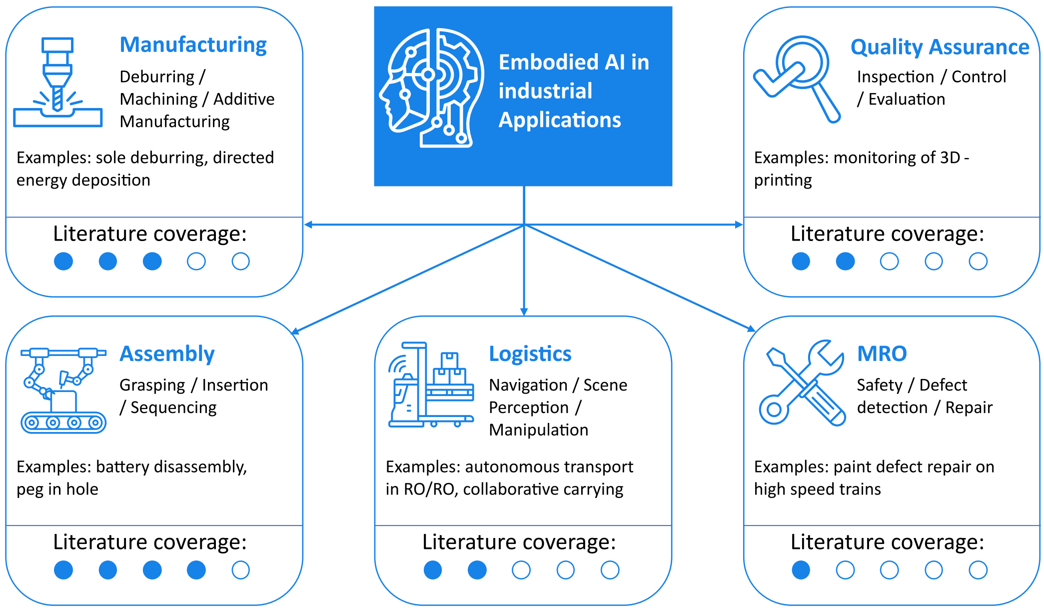

Fig. 6. Application areas of EAI in industrial environments.

A second recurring topic is HRC. Several works use assembly scenarios to study the transition from reactive assistance toward proactive collaborative behavior, including intent anticipation, real-time adaptation to operator actions, and context-sensitive support during multi-step tasks [61,63,82]. This is particularly relevant for industrial practice, where many assembly processes remain only partially automatable and depend on flexible cooperation between human workers and robotic systems [61].

Another important aspect is decision-making during execution. Instead of treating assembly as a fixed control problem, several papers model it as a closed perception-decision-execution loop in which the current task state is continuously interpreted and translated into context-dependent actions [57,82]. This is supported by virtual planning layers for validating assembly sequences [57,64], structured scene representations for spatial relations between parts [56], and mechanisms for failure detection, verification, local replanning, or corrective action when execution deviates from the nominal sequence [66,68].

However, the reviewed literature also shows that current assembly use cases are often simplified. Many studies rely on toy examples such as LEGO-like setups [65,68], peg-in-hole benchmarks [69,81], or stylized manipulation tasks that abstract away much of the variability and physical complexity of real industrial assembly [71]. Others simplify the environment through reliable pose information, CAD-based prior knowledge [56], abstracted contact dynamics [69], or a focus on high-level reasoning while fine manipulation and process-critical execution remain only partially addressed [52,82]. Thus, assembly is a justified entry point for EAI in manufacturing, but current demonstrations still capture only part of the complexity of real industrial assembly systems.

Logistics and intralogistics. Logistics and intralogistics combine navigation, scene perception, and manipulation, making them a natural end-to-end problem class for EAI (localize → plan → move → act) [76,77,83]. In the reviewed literature, the focus is less on isolated manipulation skills and more on coordinating perception, decision making, and motion in distributed, dynamic, and safety–critical environments [59,77,83].

A recurring theme is the use of multi-agent systems for dynamic self-organization, flexible role allocation, and task-dependent cooperation instead of predefined agent responsibilities [77,78]. Related work addresses collaborative transport, where robots assist humans in carrying large or heavy objects. Here, EAI supports not only motion generation, but also interaction-aware coordination, including implicit leader–follower switching and reduced physical effort for human workers [84].

Safety is central across both indoor and outdoor settings. Industrial logistics requires rapid replanning and responsive control to satisfy operational safety constraints in environments shared with humans, vehicles, or multiple robots [59,76,83]. Consequently, real-time optimization and low-latency reaction capabilities are key technical concerns in current work [59].

Much of the literature relies on RL and simulation-based development [59,77,78]. Results in multi-agent coordination, humanoid transport, and motion generation are often validated primarily in simulation [59,77,78]. This highlights a persistent sim-to-real gap, especially under out-of-distribution physical interactions, irregular terrain, and real-world disturbances that are difficult to model accurately [78,80,84].

Quality assurance. Quality assurance appears less often as a stand-alone inspection and more frequently as part of downstream action chains such as inspect → decide → repair [64,85]. In these settings, visual or multimodal inspection is not limited to defect detection, but also triggers planning, adaptation, or corrective intervention [64,85].

A recurring theme is the need for hybrid reasoning architectures. Several works indicate that foundation models alone are not sufficiently robust for industrial inspection, especially when decisions require process-specific expertise, external knowledge, or physical constraints [53]. Hybrid approaches therefore combine VLM- or LLM-based

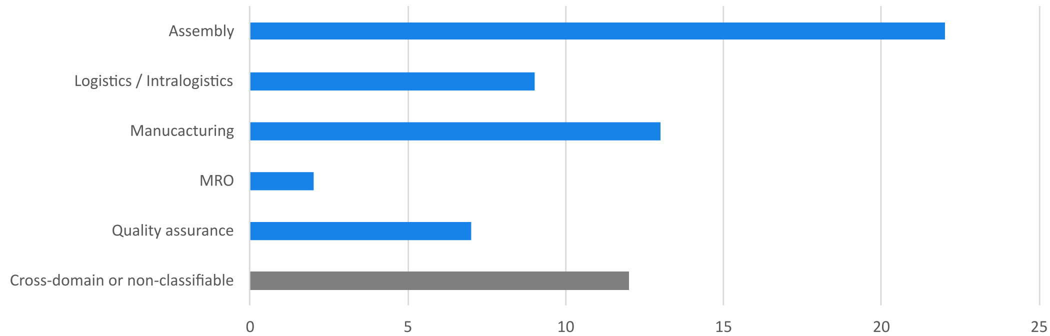

Fig. 7. Analysis of the reviewed literature based on industrial application.

perception with expert rules, structured reasoning, or physics-informed representations to improve interpretability and decision quality [53,64,85]. This is important because inspection results in manufacturing must often be translated into actionable process decisions rather than semantic descriptions alone [53,85].

The literature also shows an interest in distributed and self-correcting inspection pipelines. Multi-agent systems can support reliability through hierarchical feedback structures, where central decision layers integrate global multimodal information while local agents monitor execution and detect anomalies [77,82]. However, many pipelines still rely heavily on VLMs or LLMs. While these models are strong in semantic interpretation, they remain weaker in the quantitative regression tasks required for industrial quality assurance, such as estimating precise process parameters or millimeter-level tool paths [53,85]. This limits the transition from general-purpose reasoning to high-precision industrial inspection [53,70].

A further challenge is robustness under realistic environmental conditions. Industrial inspection remains sensitive to background noise, reflections, low-light conditions, and visually cluttered scenes [74,76]. Although adaptive perception models improve filtering and segmentation, performance can still degrade under such disturbances, leading to hesitant or overly conservative robot behavior [74,85].

Manufacturing. In the area of manufacturing and process operations, the corpus contains fewer works than in assembly, but these works are particularly close to industrial value creation because they intervene directly in process-specific operations involving tool contact, process parameters, and quality constraints [52,86,87]. Compared with assembly-oriented studies, the literature in this area is less concerned with generic manipulation and more with complete process chains that connect perception, process planning, and execution [85,87]. Typical examples include deburring, defect repair, and additive manufacturing, where the relevant challenge is not only to identify the next action, but to generate process-feasible parameters and motions under application-specific constraints [60,86,87].

At the same time, many of the reviewed contributions approach manufacturing problems through LLM- or VLM-centered pipelines [52,85]. An example of this is an LLM-centric industrial robotics framework in which agents parse natural-language manufacturing instructions into process parameters, select suitable end-effectors, and generate tool paths or robot code [52]. This is complemented by more human-centered contributions, such as intelligent assistants for operator support and workcell systems that use retrieval and natural-language interfaces to access manuals or generate robot code for non-expert users [54,61,62].

A recurring bottleneck, however, is that foundation models are typically stronger at high-level semantic planning than at the quantitative precision required for engineering tasks, such as 3D toolpath generation, process-parameter estimation, or robust execution under variable material and surface conditions [52,53,85]. As a result, this application area is highly relevant for EAI, but current work still often treats foundation models as general-purpose reasoning layers around comparatively simplified process pipelines rather than as a fully grounded basis for robust industrial execution [52,53,85].

Maintenance, Repair, and Overhaul (MRO). MRO is only weakly represented in the reviewed literature and is often operationalized through enabling capabilities such as safe navigation, defect detection, and repair rather than through fully developed MRO workflows. In this sense, the current corpus provides only limited evidence for EAI in MRO as a distinct industrial application area. Paint-defect repair is a strong example in this area, where inspection is directly linked to process decision-making and robotic repair execution rather than remaining a stand-alone diagnostic step [85]. Overall, the literature suggests that EAI in MRO is still at an early stage, with current work focusing more on prerequisite functions for maintenance assistance than on comprehensive autonomous repair processes.

Feasibility and deployment readiness. Beyond the technical capabilities discussed above, the industrial adoption of EAI ultimately depends on operational and economic feasibility. Across the reviewed corpus, however, such considerations are addressed only selectively and remain secondary to demonstrating technical functionality. A few studies report operational or economic proxy indicators such as cycle time, task completion time, fault-handling time, hardware cost, labor reduction, data-generation cost, or downtime avoidance [53,54,59,88], and low-latency or real-time responsiveness recurs as a concern in logistics settings [59]. These indicators show that feasibility is not entirely absent from the literature, but they are fragmented, application-specific, and rarely benchmarked against conventional automation. Explicit return-on-investment or payback analyses are largely absent across the corpus.

We interpret this scarcity as a finding in itself. It is consistent with the low technology-readiness observed throughout our analysis, where many contributions are validated on simplified or toy-scale tasks and evaluated primarily in laboratory or simulation settings rather than in production. As a consequence, most industrial EAI systems are currently assessed as technology-focused proof-of-concept demonstrations rather than through standardized economic or operational deployment metrics. A further deployment reality, the certification of learning-based policies for safety–critical operation, is likewise not addressed in any of the reviewed studies; we discuss the corresponding regulatory and standardization landscape separately in Section 6.

4.4. Physical embodiments

This section categorizes the reviewed literature according to the presented physical form, utilizing the previously introduced mobility-manipulation matrix (Clusters 1 to 8 in Fig. 3). The distribution (see Fig. 8) reveals a clear focus on specific hardware archetypes. In two studies, the focus is on software components independent from a specific hardware configuration (N/A). The remaining 40 studies can be grouped into three overarching mobility tiers: stationary systems,

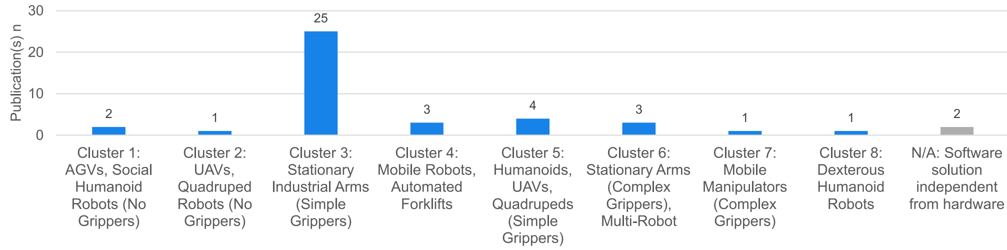

Fig. 8. Analysis results of physical embodiments in current industrial EAI research.

mobile systems for structured environments, and highly mobile systems for unstructured terrains.

Stationary systems (clusters 3 and 6). Stationary robotic systems represent the established standard in industrial automation, a trend strongly mirrored in the EAI literature. Cluster 3 (No Mobility, Low/Middle Manipulation) comprises the vast majority of the corpus (25/42 studies). This demonstrates that rather than introducing entirely new physical forms, much of the current research seeks to enhance the capabilities of already ubiquitous hardware, such as standard 6-axis arms with simple grippers, through advanced learning-based methods like RL and vision-language policies. In this cluster, nearly all researchers use cobots. Only one study relies on a traditional robotic arm within the context of a heavier repair task requiring high stiffness [85].

Only a small fraction of stationary systems push toward complex manipulation (Cluster 6, 3/42). These setups incorporate multi-fingered grippers [73,75] or dual-arm configurations with simple grippers [88] to handle more dexterous object manipulation.

Mobile systems (clusters 1, 4, and 7). Systems capable of moving across flat environments (e.g., factory floors) form the second group (6/42 studies). Cluster 1 contains two studies dealing with navigation problems of AGVs [76,89]. Clusters 4 and 7 (4/42 studies) mostly utilize mobile platforms with cobots mounted onto them. In [82,90,91], these platforms are enabling operating in a larger assembly workspace to carry out tasks on their own or in collaboration with humans. The exception to these forms is [83], which leverages a heavy-lifting device for logistic terminal operations.

Highly mobile systems (clusters 2, 5, and 8). Systems with high mobility and partly dexterous manipulation capabilities represent the upper limit of mechanical complexity in EAI, yet they remain exceptionally rare in the current literature (6/42 studies). Within these, one paper deals with a framework that can be used for different embodiments such as quadruped robots or UAVs having different manipulation skills [77], assigned to cluster 5 due to simplicity reasons. The remaining 5 studies deal with humanoid robots. Notably, the reviewed applications within these clusters do not involve the robots autonomously assembling products. Instead, their industrial applications focus on either an intralogistics context or deal with the handling/sorting of simple objects. For instance, these systems are deployed to cooperatively carry heavy objects such as boxes, stretchers, and carts alongside humans across varying terrains [84].

4.5. Coupling of learning components

An EAI system may consist of different components: Interaction, Perception, Cognition/Planning, and Action Layers (see Fig. 2). Coupling two components via differentiable operations allows gradients to flow between them, allowing continuous intermediate representations to be trained end-to-end. This typically results in the intermediate representation being non-human-readable. Conversely, decoupled components do not allow end-to-end training. As already stated in Section 2.1, coupling the information flow from input to output follows Brooks's [21] and Smith's [22] EAI hypothesis, namely, we cannot train a high-level intelligent system when it is decomposable into modular blocks that are designed such that they are human interpretable. For example, when the perception layer outputs a class value and bounding boxes for what the system sees, subsequent layers never have access to a richer representation of the system's environment. Thus, the system is somewhat bottlenecked. In contrast, modern VLMs typically follow a vision-LLM coupling strategy by linearly projecting image patch features into embeddings that are fed into the LLM.

Decoupled architectures. Completely decoupled systems are assembled using pre-trained components. Typically, a high-level planner, often in the form of an LLM, processes outputs from the perception stack and either generates code or uses API calls to perform actions. For instance, systems utilize language models to translate natural language into executable Python scripts or predefined robot action primitives, bridging the semantic-to-physical gap by dynamically adjusting to spatial coordinates provided by the perception module, e.g., [52,54,56,60,61,64,86,90]. As discussed, these systems are intrinsically limited by their architecture to reach a higher level of intelligence and transition into the field of physical AI.

Completely coupled architectures. In completely coupled systems, the previously discussed boundaries between interaction, perception, planning, and action layers dissolve. Consequently, architecture and training complexity significantly increase. These end-to-end frameworks can be roughly divided into three primary learning paradigms: large-scale supervised learning [65,80], RL [69,71,78,92], and behavior cloning via Imitation Learning (IL) [73,88]. Broadly, large-scale supervised approaches rely on massive, cross-task datasets, whereas IL frameworks typically utilize task-specific training data derived from human demonstrations. The training data requirements differ vastly between these goals. For example, training foundation VLA models for generalized capabilities requires massive, highly diverse datasets, often leveraging high-fidelity synthetic data generated automatically in physics simulators across multiple robotic embodiments [80]. Conversely, IL approaches rely on highly specific human demonstration data collected via direct teleoperation or virtual reality interfaces, prioritizing explicit motion quality and self-correction maneuvers to handle complex, task-specific dexterity [88].

In these setups, cognition and action are deeply integrated, directly mapping environmental states or raw visual inputs to low-level motor commands. Consequently, action is organically coupled within these end-to-end approaches, whereas the LLM-heavy decoupled frameworks treat action as a delayed, strictly sequential execution step rather than a continuous sensorimotor loop. On the other hand, works that demonstrate the use of end-to-end architecture typically lack an integrated interaction layer beyond simple commands and traditionally suffer from a lack of transparency. To mitigate this "black box" problem, recent integrated architectures embed Chain-of-Thought (CoT) reasoning into their policy generation, explicitly articulating intermediate

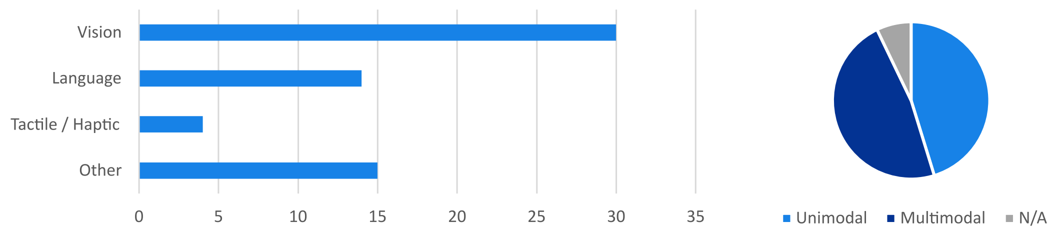

Fig. 9. Analysis of input modalities in current industrial EAI research. Papers with multimodal inputs are counted toward all included input modalities.

reasoning, subgoals, and verification criteria alongside low-level motor outputs, thereby maintaining continuous end-to-end execution while regaining human interpretability [65].

Partially coupled architectures. Partially coupled systems can move one step further into the EAI hypothesis compared to fully decoupled approaches when coupling, e.g., perception and cognition/planning, or cognition/planning with the action layer. However, architectures are more diverse in what is coupled and what is decoupled. Besides, training data is different, especially when targeting a domain-specific task. In multiple works, the interaction layer is decoupled from the rest. For example, multiple works use a pre-trained voice-to-text model to let the user interact with the EAI system, e.g., [60,66]. On the other hand, one can discuss whether a coupled interaction layer is necessary to meet the EAI hypothesis. Also, a large set of works in the survey lack an interaction layer [71,74] or even a perception and cognition/planning [55]. However, these studies aim at different research questions than building a holistic EAI system for industrial tasks.

The transition of EAI from laboratory environments to industrial real-world applications exposes a tension between the theoretical ideal of the EAI hypothesis, including the end-to-end character, and especially the practical constraints of training data. Methods that assemble pre-trained components are the first step from physical to EAI. The logical next step is to partially couple components to reach a higher level of intelligence based on the available training data and the needed generalizability across tasks. Collecting large-scale, annotated data is expensive and restricted by industrial confidentiality.

4.6. Input/output modalities

This subsection examines how industrial EAI systems can be characterized by both the information they receive and the form in which they produce actionable results. The input side captures how systems perceive and represent their environment, while the output side reflects the level at which learning-based components intervene in planning, control, and execution (see Fig. 10).

4.6.1. Input modalities

We categorize input modalities into the classes Vision, Language, Tactile/Haptic, Other, and Not Specified. Importantly, these categories are not mutually exclusive, since many systems are multimodal and combine multiple sensor and data sources; totals across modality groups therefore overlap.

Across the included studies, vision is the dominant input channel (30), followed by other modalities (15), language (14), and tactile/haptic input (4) (see Fig. 9). At the same time, the corpus is almost evenly split between unimodal (19) and multimodal (20) systems, indicating that industrial EAI frequently combines complementary information sources rather than relying on a single sensing channel. In three studies, the input modality remains unclear or is not specified. In the following, we briefly characterize each modality and describe the typical roles it plays in the reviewed systems.

Vision. Visual input is by far the most common modality in the reviewed literature corpus (30), underlining its central role in industrial EAI. In these systems, vision primarily serves as the basis for active perception and spatial reasoning. Cameras provide access to object positions, scene structure, and execution progress, making visual input the main source for grounding actions in the physical workspace [56]. Beyond classical perception tasks such as detection, localization, and inspection, vision is also used for process monitoring and, increasingly, for the construction of richer internal world representations over longer time horizons [58,72].

Beyond HRC, visual perception supports contextual awareness and human-aware coordination [64]. More generally, vision often acts as the interface through which embodied systems detect deviations between the expected and actual state of the environment, making it a key enabler of autonomous failure recovery [68]. Identifying misalignments, errors, or unexpected events is therefore a prerequisite for replanning and corrective action.

At the architectural level, the literature shows two broad tendencies. One consists of modular pipelines in which visual models generate intermediate outputs (decoupled architecture (Section 4.5)), such as object coordinates, defect parameters, or scene abstractions, which are passed to downstream planners, policies, or language-based reasoning modules [56,57,64]. While this structure aligns well with established industrial engineering practice, it often introduces a bottleneck: high-dimensional visual observations are reduced to a small set of task variables, which simplifies downstream control but may also discard relevant context. The second tendency (partially (Section 4.5) or completely coupled (Section 4.5) architecture) comprises more tightly integrated approaches in which visual representations are fed directly into learned policies, latent-space controllers, or world models [72,81,87]. These approaches aim for stronger sensorimotor coupling, but typically require more data and offer less transparency in deployment.

Language. Language-based input (text, prompts, documents, and voice) is present in 14/42 studies, but only rarely as a standalone modality. In most of the reviewed systems, language does not operate at the level of direct control. Instead, it serves as a high-level interface for task specification, planning, and coordination. Its primary function is to translate human intent into executable structure. This is particularly evident in frameworks for long-horizon industrial tasks, where language is used to decompose complex goals into subtasks, semantic plans, or other structured intermediate representations that can be passed to downstream perception, planning, or execution modules [56,57,65].

A second recurring role of language is orchestration. Many systems use LLMs, VLMs, or VLA-style components to coordinate existing perception models, planners, tool libraries, or robot skills, rather than replacing them with a single end-to-end architecture [57,64,82]. This architectural pattern aligns well with industrial requirements for modularity and reuse. At the same time, it means that language often remains supervisory in nature: it structures decisions, recovery strategies, and task progression, while execution is still carried out by downstream controllers or predefined skills. As a result, the contribution of language in many of these systems lies less in a deeply language-grounded embodied policy than in a layered composition of foundation-model reasoning on top of existing robotic pipelines.

Language also plays an important role in lowering the barrier to system configuration and adaptation. Several studies use natural language as the basis for generating code, task logic, or control specifications, including robot programs, Python-based routines, and machine-level outputs such as G-code [52,54]. Closely related to this is the emergence of low-code or no-code interaction paradigms, in which users specify machining, handling, or assembly objectives through text or voice rather than through conventional robot-programming interfaces [54,60,70].

Finally, language increasingly serves as a medium for grounding decisions in external expert knowledge. Through retrieval-augmented or tool-augmented designs, embodied systems can access technical manuals, process documentation, or troubleshooting information and incorporate these sources into planning and reasoning [53,54,82]. This is particularly relevant in industrial environments, where correct behavior often depends less on general commonsense reasoning than on adherence to domain-specific constraints, procedures, and safety rules.

Tactile/haptic. Tactile or haptic input is rare in the current literature (4/42), although contact and force information are inherently highly relevant for many industrial manipulation and assembly processes. The few available examples focus primarily on (i) force/torque sensing in manipulation and disassembly [70,91] and (ii) force/pressure information in humanoid motion and cooperative tasks [59,84]. Across these studies, tactile input appears less as a standalone modality than as a complement to vision, proprioception, or control [70,84,91]. Its main role is to capture interaction dynamics that cannot be reliably inferred from vision alone, especially in contact-rich tasks such as insertion, assembly, carrying, or disassembly [70,84,91]. In this respect, haptic information is closely tied to physical execution, as it provides direct feedback on contact events, alignment errors, load transfer, and task progress at the moment of interaction. Haptic signals are used not only for safety interlocks, but increasingly also for behavior adaptation and policy refinement [70,84]. At the same time, some systems infer interaction forces only implicitly through joint-state deviations or related internal signals rather than through dedicated external haptic sensors [84]. This broadens the functional role of haptic information, but also blurs the boundary between tactile sensing, force estimation, and proprioceptive observation.

Other. The category “Other” (15/42) comprises structured, mostly low-dimensional inputs that do not fall under classical vision, language, or tactile channels. These inputs include proprioceptive robot states, process-specific sensor data, auxiliary spatial sensors, and other task-dependent observations [76,84,87].

A central subgroup consists of proprioceptive and state-based inputs, such as joint angles, joint velocities, gripper states, poses, or related internal robot variables [72,73,84]. These signals are particularly common in reinforcement-learning-based and policy-learning settings, where they provide a compact representation of the robot’s current configuration and, in some cases, also serve as an implicit proxy for environmental interaction dynamics [72,73,78,84]. The literature shows that such internal observations are often used to stabilize control and to complement more uncertain exteroceptive inputs [72,76].

A second important subgroup consists of process-specific and industrial sensor data. In these cases, the learning-based component is not primarily driven by scene perception, but by measurements tied closely to the physics of the underlying operation, such as thermal variables, airflow, melt-pool characteristics, or other industrial sensor data, e.g., in [87]. Related to this are auxiliary spatial sensors used to improve localization and navigation, such as wheel encoders, QR-based positioning targets, or LiDAR data, which extend perception beyond standard camera-based sensing [76,77].

This category also includes more specialized observations that reflect the diversity of industrial EAI. These range from human-centered biosignals and motion-tracking data to relational or adjacency-aware state representations in multi-object or collaborative settings [63,71,78].

4.6.2. Output

Output modalities in industrial EAI are more heterogeneous than input modalities, because they reflect not only what a system perceives, but also the level at which the learning-based component intervenes in the control stack [52,57]. Across the reviewed literature, outputs range from direct physical actuation signals to geometric action parameters, symbolic task representations, and machine-executable code [52,55,57,85]. In this sense, the output side is best understood less as a set of mutually exclusive categories than as a spectrum of abstraction levels, spanning low-level control, mid-level motion specification, and high-level semantic or programmatic decision making.

At the lowest level, some systems output continuous control variables that are mapped directly to robot motion or actuation. These include joint commands, motor actions, force- or torque-related signals, and other physically grounded control variables, especially in reinforcement-learning-based or imitation-learning-based settings [55,75,84]. Such outputs represent the strongest form of embodiment, since the learned component is tightly coupled to the robot’s physical action. At the same time, this level of control places high demands on robustness, verification, and safety, which helps explain why it is currently more common in locomotion, humanoid control, or simulation-heavy research settings than in heavily constrained industrial deployment scenarios [55,59]. A second level consists of geometric or action-oriented outputs such as grasp poses, end-effector targets, trajectories, navigation commands, or other structured motion parameters [74,76]. Here, the learning-based system does not necessarily control the robot directly, but instead produces task-relevant motion specifications that can be passed to downstream planners or conventional controllers. A third level comprises symbolic or programmatic outputs [57]. In such systems, the learning-based component generates semantic task plans, skill sequences, parameter recommendations, or even machine-executable instructions such as robot code or G-code [52,85]. This type of output is especially characteristic of language-enabled and multimodal industrial systems, where the model acts less as a direct controller and more as a planner, coordinator, or code-generation interface [57]. Finally, some systems produce intermediate representations that are not final actions in themselves, but still function as operational outputs for downstream decision making [56,66,85]. These include scene graphs, defect geometries, interaction models, or other structured representations that summarize the current state of the task in a form that can be consumed by planners, policies, or execution modules.

Taken together, the literature suggests that the key distinction on the output side is the layer of abstraction at which a system produces actionable information. While some approaches operate directly at the actuator or joint level, many industrially oriented systems place learning-based outputs at a higher level, for example, as poses, semantic actions, or machine-readable plans. This indicates that, in industrial EAI, output design is strongly shaped by the need to balance embodiment with controllability, interpretability, and safety. However, the current state of the art does not yet provide end-to-end systems that satisfy these requirements with sufficient reliability for industrial deployment.

4.7. Simulation environments

Simulation and reality. Across the reviewed corpus, 25 papers deal with real-world setups, 6 contributions combine simulation and real-world experiments, and 7 records are simulation-only. For four papers, this segregation is not applicable or insufficiently specified. This distribution suggests that the literature remains primarily anchored in

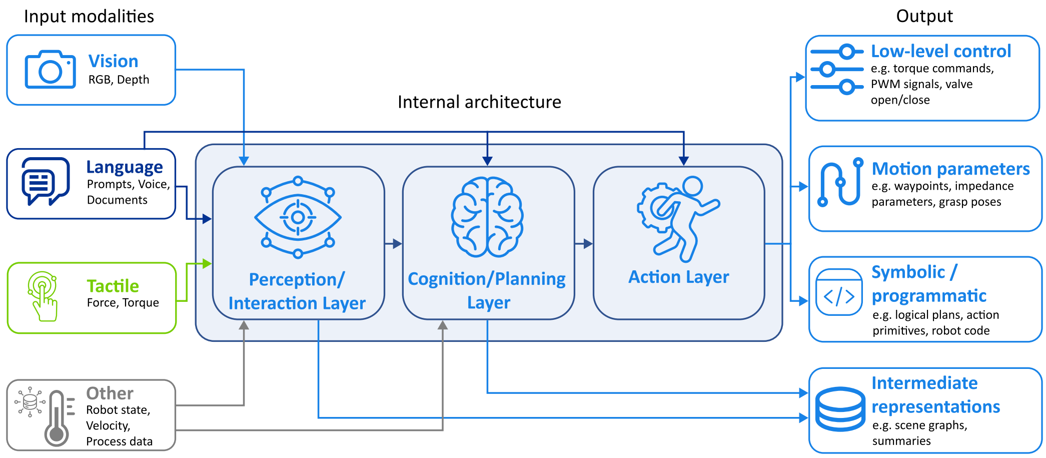

Fig. 10. Input and output modalities in the reviewed literature.

real-world validation, while simulation is used more selectively and usualy for clearly defined methodological purposes. By contrast, the sim+real and simulation-only subsets use simulation more directly as part of the methodological core, especially when repeated interaction, safe exploration, or scalable data generation is required. From the perspective of this review, the key distinction is therefore not only where experiments are conducted, but why simulation is used.

In the sim+real studies, simulation typically serves to reduce development risk, pre-train controllers, align models, or stress-test policies before deployment on real hardware, e.g., in [55,81].

In the simulation-only studies, by contrast, the simulator often functions as the primary substrate for learning and evaluation, especially in RL or task generation settings where large numbers of safe and repeatable interactions are essential [58,70–72,78]. Overall, the literature suggests that simulation in industrial EAI is not used as a default condition, but rather as a targeted methodological resource when real-world experimentation alone is insufficient.

Types of simulation and their function. Across the 13 papers that involve simulation in some form, the dominant pattern is physics-based 3D robotic simulation (including CoppeliaSim (RLBench), PyBullet, Isaac Gym/Isaac Sim, and RaiSim). Although the specific environments differ, they share a common functional logic: they provide a controllable environment in which robot kinematics, collision and contact dynamics, scene geometry, and at least basic perception mechanisms can be coupled during training or evaluation [55,58,70–72,78,81].

Within this group, simulation serves three main functions. First, it enables sample-intensive policy learning under safe and repeatable conditions, particularly for RL and world-model approaches [58,70,72,78]. Second, it supports sim-to-real workflows in which controllers or learned models are tuned in simulation before or alongside deployment [55,81]. Third, it can serve as infrastructure for scalable synthetic data generation rather than as the environment of a single learning agent [80].

A conceptually distinct use of simulation appears in the digital-twin approach of the HVAC study [62]. Here, simulation is not primarily used as a robotics sandbox for offline learning, but as a plant-specific virtual counterpart that runs alongside the physical system and supports prediction, scenario testing, and control optimization. This expands the notion of simulation in industrial EAI beyond robot training alone and highlights its role as a cyber–physical tool for process-level adaptation. At the same time, the simulation-involving subset remains relatively small and methodologically heterogeneous.

4.8. From scratch to VLAs: Learning components and training

To give an overview of the different strategies for incorporating learnable components, we categorize the literature by architectures and (additional) training strategies. Specifically, we divide the approaches into four categories based on their reliance on foundation models: (1) no foundation models, (2) LLMs, (3) VLMs, and (4) VLAs. For frameworks that use foundation models, we further distinguish between training-free deployments and those that require additional training, where applicable.

Without foundation models. The first category encompasses approaches that do not rely on foundation models. In these works, high-level interaction and broad visual reasoning are largely constrained or omitted entirely. This results from the fact that currently, only pre-trained foundation models enable off-the-shelf use or are a starting point for larger generalization and generalized reasoning. For example, [92] trains different policy algorithms from scratch for basic brick-like pick-and-place assembly. Similarly, [78] explores an approach to train agents to control bipedal robots that collaboratively handle objects. [71] trains an RL agent from scratch for compact multi-object placement without relying on explicit visual target poses. Exploring behavioral cloning, [84] utilizes a teacher-student distillation framework to train a proprioception-only policy for human-humanoid collaborative carrying. Overall, these approaches are typically highly specialized rather than generalist EAI systems for industrial tasks; they focus on exploring and enabling novel morphological capabilities or techniques for basic isolated industrial tasks.

Large Language Models (LLMs). As outlined in previous sections, LLM-heavy frameworks utilize text-based foundation models primarily for logical reasoning, task allocation, or code generation. Many of these works rely on training-free adaptation methods, such as zero-shot or in-context learning. For instance, [52] uses GPT-4 via few-shot prompting to autonomously design tool paths and act as a high-level task planner. Further, [54] also uses GPT-4 coupled with a RAG pipeline to build a workcell assistant. [90] uses GPT-4 to interpret tasks via CoT reasoning for HRC. Conversely, other approaches introduce additional training. [85] fine-tunes a small LLM model using QLoRA. Similarly, [77] adapts a pre-trained LLM through domain-specific fine-tuning on multi-agent coordination dialogues and logistics protocols.

Vision-Language Models (VLMs). Frameworks in the third category utilize VLMs to bridge semantic understanding with visual input, though they do not natively output continuous physical control actions. Directly using pre-trained VLMs without fine-tuning is demonstrated, for example, by [68], leveraging VLMs as motion failure detectors via optimized prompts. [67] uses a frozen VLM core augmented with a prompt synthesis module, extending the model's capabilities to 3D spatial reasoning. [56] utilizes pre-trained VLMs (as agents) for zero-shot reasoning to generate semantic scene graphs, enhancing robotic task allocation. [57] fine-tunes VLMs to achieve robust situational awareness in narrow-space assembly scenarios. [77] fine-tunes an ImageNet-pretrained VLM on domain-specific logistics datasets to enable domain-specific recognition of delivery objects.

Vision-Language-Action models (VLAs). The fourth category encompasses VLAs, which build upon multimodal foundation models but are designed to natively output low-level robotic actions based on integrated vision and language inputs. For instance, [65] uses the Janus-Pro VLM as a foundation and fine-tunes it on self-collected and simulation-synthesized manual data, updating the action expert parameters via IL on over 400K real-world trajectories. Further, [66] fine-tunes a LLaMA-2 backbone extended with an action head, training the system on human demonstration videos for multi-arm manipulation.

Reviewing these four categories reveals a clear trend in their design. Works that focus on novel morphologies or RL methods for explorative demonstrations of basic tasks do not build upon large-scale pre-trained foundation models. Among those that do use foundation models, the majority utilize LLMs to connect interaction inputs to action outputs. Fewer works use VLMs, and even fewer employ VLAs. This disparity likely results from the severe data bottleneck. While internet-scale text and image data are abundant for pre-training LLMs and VLMs, standard web data is insufficient for VLAs. The collection of internet-scale, generalized action/kinematic data remains an open challenge, meaning VLAs still heavily rely on expensive, task-specific, and domain-specific training data. Where research on training generalist VLA depends on new large-scale datasets like [74,80] (included in our SLR), datasets for fine-tuning industrial-domain VLAs are rare [66,79].

5. Summary

This section synthesizes the current state of EAI in industrial systems by projecting our core findings onto our guiding research questions.

Characterization and application of the term EAI (RQ1). Our conceptual analysis reveals an inconsistency within the industrial literature regarding the use of the term EAI. While the surveyed works fulfill the baseline requirement of our definition (Section 2.1), utilizing a physical or simulated embodiment, the strict theoretical characteristics of EAI, such as end-to-end sensorimotor coupling, situated intelligence, and learning constrained by morphology, are frequently absent. Admittedly, the practical necessity of profound situated intelligence is debatable for highly constrained use cases; an agent operating within a dedicated assembly cell may not require an understanding of the entire production facility. Nevertheless, incorporating broader environmental stimuli remains essential to elevate the system's overall intelligence, which is a prerequisite for solving complex, unstructured tasks with the adaptability of a human factory worker. Often, the term EAI is adopted merely as a generic label for modular systems driven by decoupled foundation models, treating physical action as a separate, sequential output rather than a deeply integrated sensorimotor process. This exposes a critical gap between the theoretical EAI definition and its current practical implementation in industrial scopes.

Research focuses, applications, and embodiments (RQ2 & RQ3). The technological trajectory of industrial EAI is driven by six core research clusters that are applied unevenly across industrial domains and heavily anchored to established hardware. *Assembly* and *Manufacturing* serve as the primary test beds for clusters focusing on autonomous task execution (Cluster 3), proactive HRC (Cluster 1), and robotic dexterity (Cluster 4). Yet, the actual operations within these domains frequently rely on simple, abstracted sequences. Alternatively, they target highly specific, unique niche applications that previously lacked extensive research but now gain a significant boost from learning-based adaptability. To enable these use cases, researchers heavily integrate foundation models for high-level reasoning and planning (Cluster 2). However, these ambitious cognitive goals are almost exclusively implemented on stationary six-axis collaborative robot arms. Furthermore, *Manufacturing* and *Quality Assurance* applications remain bottlenecked because current foundation models lack the spatial accuracy and deterministic reliability necessary to meet strict industrial tolerances. Conversely, the cluster dedicated to mobility and multi-agent systems (Cluster 5) naturally aligns with *Logistics* applications, representing one of the few areas exploring highly mobile morphologies, such as automated platforms and humanoid robots. Comprehensive *MRO* workflows, meanwhile, remain exceptionally scarce.

A persistent challenge across all these domains is the sim-to-real gap. Whether attempting to scale abstracted assembly tasks (e.g., toy examples or peg-in-hole) or deploying mobile agents into dynamic logistics facilities, real-world physical complexity remains a severe bottleneck. To directly address this, a dedicated research cluster focuses on generating specialized datasets, synthetic data, and simulations (Cluster 6). This reveals that current industrial EAI prioritizes augmenting the cognitive capabilities of standard automation hardware over deploying fundamentally new physical forms.

Architectures, modalities, and training paradigms (RQ4 & RQ5). The underlying architectures of industrial EAI expose a fundamental tension between theoretical fully integrated ideals and practical deployment constraints. While vision serves as the dominant input modality, frequently combined with language for supervisory orchestration, tactile and haptic feedback remain underutilized despite their relevance for physical manufacturing. On the output side, systems mostly favor advanced semantic plans, geometric poses, or executable code over direct actuator control to maintain interpretability and safety. This drives a prevalence of decoupled architectures, where pretrained foundation models act as modular planners rather than deeply integrated sensorimotor policies. Consequently, the training landscape is heavily skewed toward the use of standard or specifically adapted LLMs and VLMs. VLA models, which natively couple perception and physical action, remain remarkably rare due to a severe bottleneck in domain-specific training data. To mitigate this data scarcity and support highly iterative RL or IL, researchers rely on physics-based 3D simulations. However, rather than representing comprehensive operational environments, simulators are primarily deployed as targeted methodological tools for safe policy pretraining and synthetic data generation.

6. Conclusion

This SLR examined the current state of EAI in industrial operations. The reviewed literature shows that EAI has clear potential for manufacturing, assembly, logistics, quality assurance, and maintenance-related tasks, particularly where conventional automation struggles with variability, uncertainty, and flexible human–robot interaction. However, industrial EAI is still at an early stage of maturity. Many works use the terminology of EAI, but only partially realize its core principle: intelligence emerging from the coupling of embodiment, learning, perception, action, and situated interaction with the environment.

The main barriers to large-scale deployment are architectural and practical. Current systems often rely on modular foundation model pipelines that support semantic reasoning and task planning, but remain weakly grounded in physical execution. This limits robustness in contact-rich, safety–critical, and precision-demanding settings. Moreover, many demonstrations are evaluated in simplified tasks, controlled environments, or simulations, leaving open whether they can handle real industrial variability such as tolerances, disturbances, changing materials, or rare failure cases. The limited use of tactile, force, and proprioceptive feedback further restricts reliable manipulation, while missing benchmarks and validation standards make deployment readiness difficult to assess.

Certification is a further, currently unaddressed barrier: none of the reviewed studies consider how learning-based policies would be certified for safety–critical operation, even though industrial deployment must satisfy an increasingly explicit regulatory landscape. Robot safety is governed by standards such as ISO 10218-1/-2 [93,94] (with the 2025 revision integrating the collaborative-workspace and contact-force requirements formerly in ISO/TS 15066 [95]) and functional-safety standards like ISO 13849-1 [96] and IEC 62061 [97]. On the regulatory side, the EU Machinery Regulation (EU) 2023/1230 [98], applicable from January 2027, for the first time explicitly regulates safety components with fully or partially self-evolving machine-learning behavior, requiring third-party conformity assessment and a bounded task and movement space, while the EU AI Act (EU) 2024/1689 [99] classifies AI used as a safety component of machinery as high-risk. The non-determinism and limited explainability of end-to-end learned policies make conventional verification difficult, further motivating the modular, human-supervised architectures observed in our corpus.

Progress in the near future will likely come from task-specific EAI systems rather than fully autonomous general-purpose robots. Promising directions include partially coupled architectures, richer multimodal sensorimotor feedback, and high-fidelity simulation for development, and industrial benchmarks that evaluate precision, robustness, safety, recovery behavior, and economic value. This also applies to industrial feasibility: while a few selected studies report proxy indicators such as cycle time, task completion time, hardware cost, data cost, or downtime reduction, explicit ROI or payback analyses remain largely absent, highlighting the need for future benchmarks to include economic and operational metrics alongside technical performance. Thus, EAI can move from isolated demonstrations toward scalable industrial deployment by closing the gap between high-level reasoning and physically reliable execution.

CRediT authorship contribution statement

Julian Koch: Writing – review & editing, Writing – original draft, Visualization, Validation, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Philipp Prünte:** Writing – review & editing, Writing – original draft, Visualization, Validation, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Keno Moenck:** Writing – review & editing, Writing – original draft, Visualization, Validation, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Jonathan Determann:** Writing – review & editing, Writing – original draft, Visualization, Validation, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Martin Gomse:** Writing – review & editing, Supervision, Resources, Project administration. **Thorsten Schüppstuhl:** Writing – review & editing, Supervision, Resources, Project administration.

Declaration of generative AI and AI-assisted technologies in the manuscript preparation process

During the preparation of this work, the authors used generative AI tools in order to research, draft, write, summarize, and modify images (Gemini 3 for writing/Nano Banana Pro for images). The SLR itself, including all PRISMA steps, was conducted manually. After using these tools, the authors reviewed and edited the content as needed and take full responsibility for the content of the published article.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendix

A.1. Search methodology details

This section details the exact search terms and details used for gathering the studies from Scopus, IEEE Xplore, and arXiv as input for step 1 of the PRISMA workflow, see Fig. 4.

For IEEE Xplore, we used the basic search with a custom search string. For SCOPUS, we used the advanced search functionality. For retrieving arXiv records, we employed a modified version of the `arxivcollector` repository taken from https://github.com/koenraijer/arxivcollector as the official API did not reliably return results. All search engines are configured to yield records with a publication date as specified in Table 4. For records on arXiv, we defined the v1 submission date as the relevant timestamp and have modified the retrieval script accordingly.

For the searches in IEEE Xplore and arXiv, we used all available fields; for the SCOPUS search, we constrained the search to the title, abstract, and keyword fields.

IEEE Xplore: Basic search with the exact search string given in Section 3.3.2, then manually constraining the date range and downloading the results.

SCOPUS: Advanced search constraint to search only the title, abstract, and keywords fields with a date range included in the query command:

```
("Manufacturing" OR "Assembly" OR "Logistic*" OR "Production
System*" OR "MRO" OR "Maintenance, Repair and Overhaul" OR
"Maintenance, Repair, Overhaul") AND ("Embodied AI" OR "Embodied
Intelligence" OR "Agentic AI" OR "Intelligent Robotics" OR
"Physical AI" OR "General Artificial Intelligence" OR "Embodied
AGI" OR "World Models" OR "Vision?Language?Action" OR "Humanoid*")
```

arXiv: The search URL was constructed by the `arxivcollector` script. We searched all arXiv subjects and constrained the date range. As arXiv does not provide citation counts, we retrieved the citation count using the Semantic Scholar API. For arXiv, we removed wildcards and use only the following terms:

Population/Problem: Manufacturing; Assembly; Logistic; Logistics; Production System; Production Systems; MRO; Maintenance, Repair and Overhaul; Maintenance, Repair, Overhaul.

Intervention: Embodied AI; Embodied AI; Embodied Intelligence; Agentic AI; Intelligent Robotics; Physical AI; General Artificial Intelligence; Embodied AGI; World Models; World Model; Vision Language Action; Vision-Language-Action; Humanoid; Humanoids.

After record retrieval from the three sources, we applied a post-retrieval filter stage to enforce our exact search terms to be included in the fields title, abstract, or keywords, as we did not constrain the search in this regard when using the IEEE Xplore and the arXiv retrieval script. This also avoids the issue with results containing non-exact phrases as described in Section 3.3.2.

A.2. List of papers

See Table 5.

Table 5
Papers assessed in 6 through fulltext screening using the RA checklist and that were not previously rejected due to other reasons.

Title	Year	Ref.	RA1	RA2	RA3	RA4	Incl.
A 3D Reconstruction and Relocalization Method for Humanoid Welding Robots	2025	[100]	✓		✓	✓	
A 3D-printed Neuromorphic Humanoid Hand for Grasping Unknown Objects	2022	[101]	✓		✓	✓	
A Force-Position Hybrid Control for Full-Size Humanoid Robots Based on Adaptive Parameters	2025	[102]	✓		✓	✓	
A Fully-distributed Shape-aware Neural Controller for Modular Robots	2023	[103]		✓	✓	✓	
A Novel Paradigm of Robotic Machining towards Embodied Intelligent Manufacturing: Case Study on Paint Defect Repair	2026	[85]	✓	✓	✓	✓	✓
A Novel RGB-D Cross-Background Robot Grasp Detection Dataset and Background-Adaptive Grasping Network	2024	[74]	✓	✓	✓	✓	✓
A Planar Compound Eye Based Microsystem for High Precision 3D Perception	2024	[104]				✓	
A Real-Time Perception–Motion Codesign Method for Image-Based Visual Servoing in Embodied Intelligence Systems	2025	[105]	✓		✓	✓	
A Self-Adaptive High Precision Gripper for Shape Variant Components: Towards Higher Reliability and Efficiency of a Cobotic Cell	2023	[106]	✓		✓	✓	
Adaptive Mechanical Metamaterials with On-Demand Binary Local Modulus for Embodied Intelligence	2025	[107]				✓	
Adaptive Robust Control for Fuzzy Mechanical Systems in Confined Spaces: Nash Game Optimization Design	2024	[108]	✓		✓	✓	
Advanced Structural Design for Humanoid Robot Upper Limbs	2025	[109]	✓		✓	✓	
Agentic AI for Intent-Based Industrial Automation	2025	[110]	✓	✓		✓	
Agentic AI for Smart Manufacturing	2025	[111]	✓	✓		✓	
Agentic AI in Smart Manufacturing: Enabling Human-Centric Predictive Maintenance Ecosystems	2025	[112]	✓	✓		✓	
An Integrated Calibration and 3D Reconstruction Method for Humanoid Welding Robots	2025	[113]	✓		✓	✓	
An Intelligent Robotics Modular Architecture for Easy Adaptation to Novel Tasks and Applications	2023	[114]	✓		✓	✓	
Assessing Perceived Discomfort and Proxemic Behavior towards Robots: A Comparative Study between Real and Augmented Reality Presentations	2023	[115]	✓		✓	✓	
Automating Robot Failure Recovery Using Vision-Language Models With Optimized Prompts	2024	[68]	✓	✓	✓	✓	✓
Autonomous Assembly Method of 3-Arm Robot to Fix the Multipin and Hole Load Plate on a Space Station	2021	[116]	✓		✓	✓	
AutoPrune: Each Complexity Deserves a Pruning Policy	2025	[117]		✓		✓	
Balance Control of a Wheeled Humanoid Robot During the Standing-Up Process Using Backpropagation Neural Networks	2025	[118]			✓	✓	
Behavioral Planning and Parameter Meta Learning for Embodied Intelligence Robots in Adaptive Assembly	2026	[70]	✓	✓	✓	✓	✓
Behaviour Range Optimization of the Dexterous Robotic Open Parametric Hand	2025	[119]	✓		✓	✓	
BookBot: A Robotic Manipulation Benchmark for Voice-Driven Book Recognition and Grasping in Cluttered Environments	2025	[120]			✓	✓	
Category-Controllable and High-Fidelity 3D Defect Synthesis for Embodied Intelligence-based Industrial Inspection	2026	[121]	✓	✓		✓	
Coaxial Graphene/MXene Microfibers with Interfacial Buffer-Based Lightweight Distance Sensors Assisting Lossless Grasping of Fragile and Deformable Objects	2023	[122]					
Collaboration Between Industrial, Collaborative, Humanoid Robots and Humans.	2025	[123]					
Compact Multi-Object Placement Using Adjacency-Aware Reinforcement Learning	2024	[71]	✓	✓	✓	✓	✓
Composite Whole-Body Control of Two-Wheeled Robots	2025	[124]	✓		✓	✓	
Conceptual Framework Toward Embodied Collective Adaptive Intelligence	2025	[125]					
Cooperative Dual-Actor Proximal Policy Optimization Algorithm for Multi-Robot Complex Control Task	2025	[78]	✓	✓	✓	✓	✓
Cooperative Motion Planning for Mobile Humanoid Robots Based on the Manipulability Map	2024	[126]	✓		✓	✓	
Customize-Your-Joy Hand: A User-Oriented, Cost-Effective 22-DOF Platform for Future Human-Robot Community	2025	[127]			✓	✓	
Design and Evaluation of a Rigid-Soft Coupling Anthropomorphic Hand System for Grasping Performance Enhancement	2025	[128]	✓		✓	✓	
Design and Validation of a Push-Latch Gripper Made in Additive Manufacturing	2023	[129]	✓		✓	✓	
Design and Validation of a Tendon-Driven Hybrid Rigid-Soft Three-Finger Gripper for Humanoid Robots	2025	[130]	✓		✓	✓	
Design Optimizer for Planar Soft-Growing Robot Manipulators	2024	[131]				✓	
Development of a Human-Centric Autonomous Heating, Ventilation, and Air Conditioning Control System Enhanced for Industry 5.0 Chemical Fiber Manufacturing	2025	[62]	✓	✓	✓	✓	✓
Development of a Simple and Novel Digital Twin Framework for Industrial Robots in Intelligent Robotics Manufacturing	2024	[132]	✓		✓	✓	
DreamArrangement: Learning Language-Conditioned Robotic Rearrangement of Objects via Denoising Diffusion and VLM Planner	2025	[133]		✓	✓	✓	
Dual-Arm Mobile Manipulation Planning of a Long Deformable Object in Industrial Installation	2023	[134]	✓		✓	✓	
Dual-Arm Trajectory Tracking Algorithm for Humanoid Robot Based on Fully-Actuated System Approach	2025	[135]			✓	✓	
Dynamic Affect-Based Motion Planning of a Humanoid Robot for Human-Robot Collaborative Assembly in Manufacturing	2024	[136]	✓		✓	✓	
Dynamic Torso Compliance Control for Standing and Walking Balance of Position-Controlled Humanoid Robots	2021	[137]	✓		✓	✓	
Echo: An Open-Source, Low-Cost Teleoperation System with Force Feedback for Dataset Collection in Robot Learning	2025	[138]	✓		✓	✓	
Effects of Robotic Expertise and Task Knowledge on Physical Ergonomics and Joint Efficiency in a Human-Robot Collaboration Task	2023	[139]	✓		✓	✓	
Embodied Intelligence: Bionic Robot Controller Integrating Environment Perception, Autonomous Planning, and Motion Control	2024	[140]	✓		✓	✓	
Embodied Intelligence in Manufacturing: Leveraging Large Language Models for Autonomous Industrial Robotics	2025	[52]	✓	✓	✓	✓	✓
Embodied Intelligence in RO/RO Logistic Terminal: Autonomous Intelligent Transportation Robot Architecture	2025	[83]	✓	✓	✓	✓	✓
Embodying Soft Robots with Octopus-Inspired Hierarchical Suction Intelligence	2025	[141]	✓		✓	✓	
Empowered Smart Manufacturing Based on an Intelligent Robotic Workcell Assistant	2025	[54]	✓	✓	✓	✓	✓
Enabling Flexibility in Manufacturing by Integrating Shopfloor and Process Perception for Mobile Robot Workers	2021	[142]	✓		✓	✓	
Enhancing Human-Robot Teleoperation through Depth Sensing and eXtended Reality	2025	[143]	✓		✓	✓	
ePA*SE: Edge-based Parallel A* for Slow Evaluations	2022	[144]	✓		✓	✓	
Exploring the Cost of Interruptions in Human-Robot Teaming	2023	[145]			✓	✓	
Field-Programmable Robotic Folding Sheet	2025	[146]			✓	✓	
FieldWorkArena: Agentic AI Benchmark for Real Field Work Tasks	2025	[147]	✓			✓	
“Follower” to “Collaborator”: A Robot Proactive Collaborative Controller Based on Human Multimodal Information for 3D Handling/Assembly Scenarios	2024	[63]	✓	✓	✓	✓	✓
Frictional and Prismatic Pin-Array Gripper for Universal Gripping and Stable Tool Manipulation	2025	[148]	✓		✓	✓	
From Insight to Autonomous Execution: VLM-enhanced Embodied Agents towards Digital Twin-Assisted Human-Robot Collaborative Assembly	2026	[57]	✓	✓	✓	✓	✓
From Observation to Action: Latent Action-based Primitive Segmentation for VLA Pre-training in Industrial Settings	2025	[79]	✓	✓	✓	✓	✓

(*continued on next page*)

Table 5 (*continued*).

Title	Year	Ref.	RA1	RA2	RA3	RA4	Incl.
From Production Logistics to Smart Manufacturing: The Vision for a New RoboCup Industrial League	2025	[149]	✓				
Future Factories with 6G: Agentic AI and Cyber-Physical Digital Twins	2025	[150]					
GEX: Democratizing Dexterity with Fully-Actuated Dexterous Hand and Exoskeleton Glove	2025	[151]	✓		✓	✓	
Graph-Fused Vision-Language-Action for Policy Reasoning in Multi-Arm Robotic Manipulation	2025	[66]	✓	✓	✓	✓	✓
GY-SLAM: A Dense Semantic SLAM System for Plant Factory Transport Robots	2024	[152]		✓	✓	✓	
HCDA: A Hidden Cross-Domain Authentication Protocol for Embodied Intelligence in Smart Manufacturing	2025	[153]	✓			✓	
Human and Humanoid-in-the-Loop (HHitL) Ecosystem: An Industry 5.0 Perspective	2025	[154]	✓	✓	✓		
Human-AI Co-Embodied Intelligence for Scientific Experimentation and Manufacturing	2025	[155]	✓	✓		✓	
Human-Like Movements of Industrial Robots Positively Impact Observer Perception	2023	[156]			✓	✓	
Humanoid Ionotronic Skin for Smart Object Recognition and Sorting	2023	[157]		✓		✓	
Humanoid Loco-Manipulations Using Combined Fast Dense 3D Tracking and SLAM With Wide-Angle Depth-Images	2024	[158]	✓		✓	✓	
Hybrid Agentic AI and Multi-Agent Systems in Smart Manufacturing	2025	[159]	✓	✓		✓	
Hybrid Reasoning for Perception, Explanation, and Autonomous Action in Manufacturing	2025	[53]	✓	✓	✓	✓	✓
Impact of Verbal Instructions and Deictic Gestures of a Cobot on the Performance of Human Coworkers	2024	[160]	✓		✓	✓	
Improving World Models for Robot Arm Grasping with Backward Dynamics Prediction	2024	[72]	✓	✓	✓	✓	✓
Industrial Collaborative Robotics, Humanoid Cobots, and Social Embodiment: A Critical Look on Human–Robot Interaction	2025	[161]	✓	✓	✓		
Industrial Foundation Model	2025	[162]	✓	✓		✓	
Integrating Virtual Twin and Deep Neural Networks for Efficient and Energy-Aware Robotic Deburring in Industry 4.0	2023	[163]	✓		✓	✓	
Integration of Robotic Vision and Automatic Tool Changer Based on Sequential Motion Primitive for Performing Assembly Tasks	2023	[164]	✓		✓	✓	
Intelligent Humanoids in Manufacturing to Address Worker Shortage and Skill Gaps: Case of Tesla Optimus	2023	[165]	✓		✓	✓	
Intelligent Robot Assistants for the Integration of Neurodiverse Operators in Manufacturing Industry	2024	[61]	✓	✓	✓	✓	✓
InternData-A1: Pioneering High-Fidelity Synthetic Data for Pre-training Generalist Policy	2025	[80]	✓	✓	✓	✓	✓
Intrinsic Motivations and Planning to Explain Tool-Use Development: A Study With a Simulated Robot Model	2022	[166]		✓	✓	✓	
ISyHand: A Dexterous Multi-Finger Robot Hand with an Articulated Palm	2025	[167]	✓		✓	✓	
JEMA-SINDYc: End-to-end Control Using Joint Embedding Multimodal Alignment in Directed Energy Deposition	2025	[87]	✓	✓	✓	✓	✓
Large Language Model for Humanoid Cognition in Proactive Human-Robot Collaboration	2024	[90]	✓	✓	✓	✓	✓
Learning Accurate and Efficient Three-Finger Grasp Generation in Clutters with an Auto-Annotated Large-Scale Dataset	2025	[168]	✓	✓	✓	✓	✓
Learning Agile Hybrid Whole-body Motor Skills for Thruster-Aided Humanoid Robots	2022	[55]	✓	✓	✓	✓	✓
Learning Dynamic Robot-to-Robot Object Handover	2023	[169]	✓		✓	✓	
Learning Human-Humanoid Coordination for Collaborative Object Carrying	2025	[84]	✓	✓	✓	✓	✓
Leveraging Procedural Generation for Learning Autonomous Peg-in-Hole Assembly in Space	2024	[69]	✓	✓	✓	✓	✓
MagicGripper: A Mini-MagicTac Integrated Gripper Enabling Multimodal Perception in Contact-Rich Manipulation	2025	[170]	✓		✓	✓	
MagicGripper: A Multimodal Sensor-Integrated Gripper for Contact-Rich Robotic Manipulation	2025	[171]	✓		✓	✓	
ManualVLA: A Unified VLA Model for Chain-of-Thought Manual Generation and Robotic Manipulation	2025	[65]	✓	✓	✓	✓	✓
MAPF-World: Action World Model for Multi-Agent Path Finding	2025	[172]		✓		✓	
Markerless Force Estimation via SuperPoint-SIFT Fusion and Finite Element Analysis: A Sensorless Solution for Deformable Object Manipulation	2025	[173]	✓			✓	
Mechanical Design of a New Anthropomorphic Robot for Fastening in Wing-Box	2021	[174]	✓		✓	✓	
Mimic-One: A Scalable Model Recipe for General Purpose Robot Dexterity	2025	[73]	✓	✓	✓	✓	✓
Mobile Robot Obstacle Detection and Avoidance with NAV-YOLO	2024	[89]	✓	✓	✓	✓	✓
Model Reconstruction-Based Optimal Adaptive Prescribed Time Control for Multi-Stage Precision Robotic Arms: Inequality Constraints, Uncertainties, Disturbances	2025	[175]			✓	✓	
MPLP: Massively Parallelized Lazy Planning	2021	[176]	✓		✓	✓	
Multimodal Perception for Indoor Mobile Robotics Navigation and Safe Manipulation	2025	[76]	✓	✓	✓	✓	✓
Multi-Modal Sensing and Perception for Humanoid Service Robots	2023	[177]			✓	✓	
Natural Entanglement Inspired Cilia-Like Soft Gripper for Rapid Adaptive Grasping	2025	[178]	✓		✓	✓	
OC-HMAS: Dynamic Self-Organization and Self-Correction in Heterogeneous Multiagent Systems Using Multimodal Large Models	2025	[77]	✓	✓	✓	✓	✓
Open Arms: Open-Source Arms, Hands & Control	2022	[179]		✓	✓	✓	
OpenPyRo-A1: An Open Python-Based Low-Cost Bimanual Robot for Embodied AI	2026	[88]	✓	✓	✓	✓	✓
Optimizing Path Planning for Automated Guided Vehicles in Constrained Warehouse Environments: Addressing the Challenges of Non-Rotary Platforms and Irregular Layouts	2025	[180]	✓		✓	✓	
PaTS-Wheel: A Passively-Transformable Single-Part Wheel for Mobile Robot Navigation on Unstructured Terrain	2024	[181]	✓		✓	✓	
Perception-Decision-Execution Coordination Mechanism Driven Dynamic Autonomous Collaboration Method for Human-like Collaborative Robot Based on Multimodal Large Language Model	2026	[82]	✓	✓	✓	✓	✓
Performance-Enhanced Flexible Self-Powered Tactile Sensor Arrays Based on Lotus Root-Derived Porous Carbon for Real-Time Human–Machine Interaction of the Robotic Snake	2024	[182]			✓	✓	
Physical AI Agents: Integrating Cognitive Intelligence with Real-World Action	2025	[183]					
Preparatory Manipulation Planning Using Automatically Determined Single and Dual Arm	2020	[184]	✓		✓	✓	
Q-Bot: Heavy Object Carriage Robot for in-House Logistics Based on Universal Vacuum Gripper	2020	[185]	✓		✓	✓	
Reinforcement Learning Control of a Humanoid Robotic Hand Actuated by Shape Memory Alloy	2021	[75]	✓	✓	✓	✓	✓
Reinforcement Learning with Vision-based State Estimation for Generalized Robotic Assembly	2025	[92]	✓	✓	✓	✓	✓
Research on Human-Computer Interaction Interface Design Based on User Experience	2024	[186]	✓		✓		
Research on Human–Machine Collaboration System for Intelligent Manufacturing - Based on Multi-source Data Fusion	2025	[187]	✓				
Research on Humanoid Robot Motion Generation and Real-Time Trajectory Optimization Based on Sparse Attention Mechanism in Deep Reinforcement Learning	2025	[59]	✓	✓	✓	✓	✓
Research on the Application of Embodied Intelligent Robots in Logistics Support	2025	[188]	✓	✓			
Residual Force Control for Agile Human Behavior Imitation and Extended Motion Synthesis	2020	[189]		✓		✓	
Reversible Execution for Robustness in Embodied AI and Industrial Robots	2021	[190]	✓		✓	✓	
Revolutionizing Battery Disassembly: The Design and Implementation of a Battery Disassembly Autonomous Mobile Manipulator Robot(BEAM-1)	2024	[91]	✓	✓	✓	✓	✓
Revolutionizing Clean Energy Labs: Robotic Imitation Learning for Efficient Fabrication AI-powered Electrical Units Assembly Platform	2025	[60]	✓	✓	✓	✓	✓

(*continued on next page*)

Table 5 (*continued*).

Title	Year	Ref.	RA1	RA2	RA3	RA4	Incl.
RL-augmented MPC Framework for Agile and Robust Bipedal Footstep Locomotion Planning and Control	2024	[191]		✓	✓	✓	
RoboHorizon: An LLM-Assisted Multi-View World Model for Long-Horizon Robotic Manipulation	2025	[58]	✓	✓	✓	✓	✓
RoboOS: A Hierarchical Embodied Framework for Cross-Embodiment and Multi-Agent Collaboration	2025	[192]		✓	✓	✓	
Robot Compliance Control Framework for Grinding Thin-Walled Parts with Unknown Surface: Deformation and Orientation Adaptation	2026	[193]	✓		✓	✓	
Robotic Automation in Apparel Manufacturing: A Novel Approach to Fabric Handling and Sewing	2025	[194]	✓		✓	✓	
SAM-RL: Sensing-aware Model-Based Reinforcement Learning via Differentiable Physics-Based Simulation and Rendering	2025	[81]	✓	✓	✓	✓	✓
Scalable Inflatable Origami Robot Modules With Proprioceptive Closed-Loop Shape Control	2025	[195]	✓		✓	✓	
Semantic 3D Scene Segmentation for Robotic Assembly Process Execution	2023	[196]	✓		✓	✓	
Semi-Autonomous, Virtual Reality Based Robotic Telemanipulation for the Execution of Peg-In-Hole Assembly Tasks	2024	[197]	✓			✓	
Single Layered Soft Skin Actuated with Twisted and Mandrel Coiled Actuators for Soft Robotics	2024	[198]				✓	
Skeletal Feature Compensation for Imitation Learning with Embodiment Mismatch	2022	[199]		✓	✓	✓	
Smart Factory Humanoid Robots with Cognitive Architectures for Human-Inspired Tasks	2025	[200]		✓	✓	✓	
Soft-Rigid Coupled Blade Leg Achieves Spatio-temporal Terrain Classification with Minimal Sensor Configuration	2025	[201]	✓		✓	✓	
Step Toward Deploying the Torque-Controlled Robot TALOS on Industrial Operations	2023	[202]	✓		✓	✓	
STORM: Segment, Track, and Object Re-Localization from a Single Image	2025	[203]	✓	✓		✓	
Terrain-Aware Hierarchical Control Framework for Dynamic Locomotion of Humanoid Robots	2025	[204]	✓		✓	✓	
Three-Dimensional-Grounded Vision-Language Framework for Robotic Task Planning: Automated Prompt Synthesis and Supervised Reasoning	2026	[67]	✓	✓	✓	✓	✓
Time-to-Fault Prediction Framework for Automated Manufacturing in Humanoid Robotics Using Deep Learning	2025	[205]	✓	✓		✓	
Towards Intelligent Robotic Sole Deburring: From Burrs Identification to Path Planning	2024	[86]	✓	✓	✓	✓	✓
Train a Robot to Climb Staircase Using Vision-Base System	2022	[206]		✓		✓	
Transporting Heavy Payloads with a Humanoid Riding a Hoverboard	2025	[207]	✓		✓	✓	
Vision Language Model-Enhanced Embodied Intelligence for Digital Twin-Assisted Human-Robot Collaborative Assembly	2025	[64]	✓	✓	✓	✓	✓
Vision-Driven Robotic Grasping Order Generation Using Segmentation and Relative Positioning in a Cluttered Environment	2025	[208]	✓	✓		✓	
VLM-MSGraph: Vision Language Model-enabled Multi-hierarchical Scene Graph for Robotic Assembly	2025	[56]	✓	✓	✓	✓	✓
Wheeled Humanoid Bilateral Teleoperation with Position-Force Control Modes for Dynamic Loco-Manipulation	2024	[209]	✓		✓	✓	

Data availability

No data was used for the research described in the article.

References

[1] P.P. Ray, ChatGPT: A comprehensive review on background, applications, key challenges, bias, ethics, limitations and future scope, Internet Things Cyber-Physical Syst. 3 (2023) 121–154, http://dx.doi.org/10.1016/j.iotcps.2023.04.003, URL https://linkinghub.elsevier.com/retrieve/pii/S266734522300024X.

[2] T. Feng, X. Wang, Y.-G. Jiang, W. Zhu, Embodied AI: From LLMs to world models, 2025, http://dx.doi.org/10.48550/ARXIV.2509.20021, Version Number: 1. URL https://arxiv.org/abs/2509.20021.

[3] D. Deming, Z. Hitzig, A. Chatterji, et al., How People Use ChatGPT, Tech. rep., OpenAI and Harvard University, 2025.

[4] M. Hachman, New report reveals what chatbots are used for, 2024, https://www.pcworld.com/article/2418551/new-report-reveals-what-chatbots-are-used-for.html.

[5] R.K. Megalingam, S.R.R. Vadivel, S.S. Kotaprolu, B. Nithul, D.V. Kumar, G. Rudravaram, Cleaning robots: A review of sensor technologies and intelligent control strategies for cleaning, J. Field Robot. 42 (5) (2025) 2234–2259, http://dx.doi.org/10.1002/rob.22515, URL https://onlinelibrary.wiley.com/doi/10.1002/rob.22515.

[6] J. Wu, R. Antonova, A. Kan, M. Lepert, A. Zeng, S. Song, J. Bohg, S. Rusinkiewicz, T. Funkhouser, TidyBot: Personalized robot assistance with large language models, 2023, http://dx.doi.org/10.48550/ARXIV.2305.05658, Version Number: 2. URL https://arxiv.org/abs/2305.05658.

[7] Y. Avigal, L. Berscheid, T. Asfour, T. Kröger, K. Goldberg, SpeedFolding: Learning efficient bimanual folding of garments, in: 2022 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), IEEE, Kyoto, Japan, 2022, pp. 1–8, http://dx.doi.org/10.1109/IROS47612.2022.9981402, URL https://ieeexplore.ieee.org/document/9981402/.

[8] A. Kelkar, C. Jansen, F. Chu, M. Patel, M. Chui, M. Robertson, R. Hess, Humanoid Robots: Crossing the Chasm from Concept to Commercial Reality, McKinsey & Compan, 2025, URL https://www.mckinsey.de/industries/industrials/our-insights/humanoid-robots-crossing-the-chasm-from-concept-to-commercial-reality. (Accessed 05 May 2026).

[9] BMW Group, First humanoid robot introduced in plant leipzig, 2026, URL https://www.bmwgroup.com/en/news/general/2026/humanoid-robot-in-leipzig.html. (Accessed 05 May 2026).

[10] Boston Dynamics, Hyundai mobis forms strategic collaboration framework with boston dynamics, 2026, URL https://bostondynamics.com/news/hyundai-mobis-forms-strategic-collaboration-framework-with-boston-dynamics/. (Accessed 05 May 2026).

[11] D. Wu, Q. Zhao, J. Fan, J. Qi, P. Zheng, J. Hu, H2R bridge: Transferring vision-language models to few-shot intention meta-perception in human robot collaboration, J. Manuf. Syst. 80 (2025) 524–535, http://dx.doi.org/10.1016/j.jmsy.2025.03.016, URL https://www.sciencedirect.com/science/article/pii/S0278612525000779.

[12] D. Wu, Q. Zhao, S. Zhang, J. Qi, J. Ma, G.-N. Zhu, J. Hu, Assembrain: Towards embodied intelligence paradigms for human–robot collaboration in manufacturing, Comput. Ind. Eng. 217 (2026) 112070, http://dx.doi.org/10.1016/j.cie.2026.112070, URL https://www.sciencedirect.com/science/article/pii/S0360835226002718.

[13] H. Liu, D. Guo, A. Cangelosi, Embodied intelligence: A synergy of morphology, action, perception and learning, ACM Comput. Surv. 57 (7) (2025) 186:1–186:36, http://dx.doi.org/10.1145/3717059, URL https://dl.acm.org/doi/10.1145/3717059.

[14] Y. Zheng, L. Yao, Y. Su, Y. Zhang, Y. Wang, S. Zhao, Y. Zhang, L.-P. Chau, A survey of embodied learning for object-centric robotic manipulation, Mach. Intell. Res. 22 (4) (2025) 588–626, http://dx.doi.org/10.1007/s11633-025-1542-8, URL https://link.springer.com/10.1007/s11633-025-1542-8.

[15] Y. Liu, W. Chen, Y. Bai, X. Liang, G. Li, W. Gao, L. Lin, Aligning cyber space with physical world: A comprehensive survey on embodied AI, IEEE/ASME Trans. Mechatronics (2025) 1–22, http://dx.doi.org/10.1109/TMECH.2025.3574943, URL https://ieeexplore.ieee.org/document/11098567.

[16] Y. Liu, Q. Sun, D.R. Kapadia, Integrating large language models into robotic autonomy: A review of motion, voice, and training pipelines, AI 6 (7) (2025) 158, http://dx.doi.org/10.3390/ai6070158, URL https://www.mdpi.com/2673-2688/6/7/158.

[17] Z. Xu, K. Wu, J. Wen, J. Li, N. Liu, Z. Che, J. Tang, A survey on robotics with foundation models: Toward embodied AI, 2024, http://dx.doi.org/10.48550/arXiv.2402.02385, arXiv:2402.02385. URL http://arxiv.org/abs/2402.02385.

[18] F. Piccialli, D. Chiaro, S. Sarwar, D. Cerciello, P. Qi, V. Mele, AgentAI: A comprehensive survey on autonomous agents in distributed AI for industry 4.0, Expert Syst. Appl. 291 (2025) 128404, http://dx.doi.org/10.1016/j.eswa.2025.128404, URL https://linkinghub.elsevier.com/retrieve/pii/S0957417425020238.

[19] J. Duan, S. Yu, H.L. Tan, H. Zhu, C. Tan, A survey of embodied AI: From simulators to research tasks, IEEE Trans. Emerg. Top. Comput. Intell. 6 (2) (2022) 230–244, http://dx.doi.org/10.1109/TETCI.2022.3141105, URL https://ieeexplore.ieee.org/document/9687596/.

[20] A.M. Turing, I.—computing machinery and intelligence, Mind LIX (236) (1950) 433–460, http://dx.doi.org/10.1093/mind/LIX.236.433.

[21] R.A. Brooks, Intelligence without representation, Artificial Intelligence 47 (1) (1991) 139–159, http://dx.doi.org/10.1016/0004-3702(91)90053-M.

[22] L. Smith, M. Gasser, The development of embodied cognition: Six lessons from babies, Artif. Life 11 (1–2) (2005) 13–29, http://dx.doi.org/10.1162/1064546053278973.

[23] N. Roy, I. Posner, T. Barfoot, P. Beaudoin, Y. Bengio, J. Bohg, O. Brock, I. Depatie, D. Fox, D. Koditschek, T. Lozano-Perez, V. Mansinghka, C. Pal, B. Richards, D. Sadigh, S. Schaal, G. Sukhatme, D. Therien, M. Toussaint, M.V. de Panne, From machine learning to robotics: Challenges and opportunities for embodied intelligence, 2021, http://dx.doi.org/10.48550/arXiv.2110.15245, arXiv:2110.15245.

[24] R. Pfeifer, F. Iida, Embodied artificial intelligence: Trends and challenges, 3139, 2003, p. 26, http://dx.doi.org/10.1007/978-3-540-27833-7_1,

[25] R. Chrisley, Embodied artificial intelligence, Artificial Intelligence 149 (1) (2003) 131–150, http://dx.doi.org/10.1016/S0004-3702(03)00055-9.

[26] T. Froese, T. Ziemke, Enactive artificial intelligence: Investigating the systemic organization of life and mind, Artificial Intelligence 173 (3) (2009) 466–500, http://dx.doi.org/10.1016/j.artint.2008.12.001.

[27] J. Duan, S. Yu, H.L. Tan, H. Zhu, C. Tan, A survey of embodied AI: From simulators to research tasks, IEEE Trans. Emerg. Top. Comput. Intell. 6 (2) (2022) 230–244, http://dx.doi.org/10.1109/TETCI.2022.3141105.

[28] Y. Ma, Z. Song, Y. Zhuang, J. Hao, I. King, A survey on vision-language-action models for embodied AI, 2025, http://dx.doi.org/10.48550/arXiv.2405.14093, arXiv:2405.14093.

[29] L. Ren, J. Dong, S. Liu, L. Zhang, L. Wang, Embodied intelligence toward future smart manufacturing in the era of AI foundation model, IEEE/ASME Trans. Mechatronics 30 (4) (2025) 2632–2642, http://dx.doi.org/10.1109/TMECH.2024.3456250.

[30] R. Chrisley, T. Ziemke, Embodiment, 2006, http://dx.doi.org/10.1002/0470018860.s00172.

[31] D.E. Koditschek, What is robotics? Why do we need it and how can we get it? Annu. Rev. Control. Robot. Auton. Syst. 4 (Volume 4, 2021) (2021) 1–33, http://dx.doi.org/10.1146/annurev-control-080320-011601.

[32] M. Shridhar, J. Thomason, D. Gordon, Y. Bisk, W. Han, R. Mottaghi, L. Zettlemoyer, D. Fox, ALFRED: A benchmark for interpreting grounded instructions for everyday tasks, 2019, http://dx.doi.org/10.48550/ARXIV.1912.01734, Version Number: 2. URL https://arxiv.org/abs/1912.01734.

[33] C. Li, R. Zhang, J. Wong, C. Gokmen, S. Srivastava, R. Martín-Martín, C. Wang, G. Levine, W. Ai, B. Martinez, H. Yin, M. Lingelbach, M. Hwang, A. Hiranaka, S. Garlanka, A. Aydin, S. Lee, J. Sun, M. Anvari, M. Sharma, D. Bansal, S. Hunter, K.-Y. Kim, A. Lou, C.R. Matthews, I. Villa-Renteria, J.H. Tang, C. Tang, F. Xia, Y. Li, S. Savarese, H. Gweon, C.K. Liu, J. Wu, L. Fei-Fei, BEHAVIOR-1K: A human-centered, embodied AI benchmark with 1,000 everyday activities and realistic simulation, 2024, http://dx.doi.org/10.48550/ARXIV.2403.09227, Version Number: 1. URL https://arxiv.org/abs/2403.09227.

[34] B. Zitkovich, T. Yu, S. Xu, P. Xu, T. Xiao, F. Xia, J. Wu, P. Wohlhart, S. Welker, A. Wahid, Q. Vuong, V. Vanhoucke, H. Tran, R. Soricut, A. Singh, J. Singh, P. Sermanet, P.R. Sanketi, G. Salazar, M.S. Ryoo, K. Reymann, K. Rao, K. Pertsch, I. Mordatch, H. Michalewski, Y. Lu, S. Levine, L. Lee, T.-W.E. Lee, I. Leal, Y. Kuang, D. Kalashnikov, R. Julian, N.J. Joshi, A. Irpan, B. Ichter, J. Hsu, A. Herzog, K. Hausman, K. Gopalakrishnan, C. Fu, P. Florence, C. Finn, K.A. Dubey, D. Driess, T. Ding, K.M. Choromanski, X. Chen, Y. Chebotar, J. Carbajal, N. Brown, A. Brohan, M.G. Arenas, K. Han, RT-2: Vision-language-action models transfer web knowledge to robotic control, in: J. Tan, M. Toussaint, K. Darvish (Eds.), Proceedings of the 7th Conference on Robot Learning, in: Proceedings of Machine Learning Research, Vol. 229, PMLR, 2023, pp. 2165–2183, URL https://proceedings.mlr.press/v229/zitkovich23a.html.

[35] M.J. Kim, K. Pertsch, S. Karamcheti, T. Xiao, A. Balakrishna, S. Nair, R. Rafailov, E. Foster, G. Lam, P. Sanketi, Q. Vuong, T. Kollar, B. Burchfiel, R. Tedrake, D. Sadigh, S. Levine, P. Liang, C. Finn, Openvla: An open-source vision-language-action model, 2024, http://dx.doi.org/10.48550/ARXIV.2406.09246, Version Number: 3. URL https://arxiv.org/abs/2406.09246.

[36] K. Black, N. Brown, D. Driess, A. Esmail, M. Equi, C. Finn, N. Fusai, L. Groom, K. Hausman, B. Ichter, S. Jakubczak, T. Jones, L. Ke, S. Levine, A. Li-Bell, M. Mothukuri, S. Nair, K. Pertsch, L.X. Shi, J. Tanner, Q. Vuong, A. Walling, H. Wang, U. Zhilinsky, π_0: A vision-language-action flow model for general robot control, 2024, http://dx.doi.org/10.48550/ARXIV.2410.24164, Version Number: 4. URL https://arxiv.org/abs/2410.24164.

[37] A.E. Adl, The Cognitive Digital Twins: Vision, Architecture Framework and Categories, Technical report, 2016, Available at: https://www.slideshare.net/aeladl/the-cognitive-digital-twin. URL https://www.slideshare.net/aeladl/the-cognitive-digital-twin.

[38] X. Zheng, X. Hu, R. Arista, J. Lu, J. Sorvari, J. Lentes, F. Ubis, D. Kiritsis, A semantic-driven tradespace framework to accelerate aircraft manufacturing system design, J. Intell. Manuf. 35 (1) (2024) 175–198, http://dx.doi.org/10.1007/s10845-022-02043-7.

[39] Y. Liu, T. Ji, X. Guo, X. Xu, J. Polzer, A survey of cognitive digital twin and the potential use of LLMs, Manuf. Lett. 44 (2025) 1242–1253, http://dx.doi.org/10.1016/j.mfglet.2025.06.144.

[40] A. Brohan, N. Brown, J. Carbajal, Y. Chebotar, J. Dabis, C. Finn, K. Gopalakrishnan, K. Hausman, A. Herzog, J. Hsu, J. Ibarz, B. Ichter, A. Irpan, T. Jackson, S. Jesmonth, N. Joshi, R. Julian, D. Kalashnikov, Y. Kuang, I. Leal, K.-H. Lee, S. Levine, Y. Lu, U. Malla, D. Manjunath, I. Mordatch, O. Nachum, C. Parada, J. Peralta, E. Perez, K. Pertsch, J. Quiambao, K. Rao, M. Ryoo, G. Salazar, P. Sanketi, K. Sayed, J. Singh, S. Sontakke, A. Stone, C. Tan, H. Tran, V. Vanhoucke, S. Vega, Q. Vuong, F. Xia, T. Xiao, P. Xu, S. Xu, T. Yu, B. Zitkovich, RT-1: Robotics transformer for real-world control at scale, in: Robotics: Science and Systems XIX, Robotics: Science and Systems Foundation, 2023, http://dx.doi.org/10.15607/RSS.2023.XIX.025, URL http://www.roboticsproceedings.org/rss19/p025.pdf.

[41] Octo Model Team, D. Ghosh, H. Walke, K. Pertsch, K. Black, O. Mees, S. Dasari, J. Hejna, T. Kreiman, C. Xu, J. Luo, Y.L. Tan, L.Y. Chen, P. Sanketi, Q. Vuong, T. Xiao, D. Sadigh, C. Finn, S. Levine, Octo: An open-source generalist robot policy, 2024, http://dx.doi.org/10.48550/ARXIV.2405.12213, Version Number: 2. URL https://arxiv.org/abs/2405.12213.

[42] Open X-embodiment: Robotic learning datasets and RT-x models, 2023, http://dx.doi.org/10.48550/ARXIV.2310.08864, URL https://arxiv.org/abs/2310.08864.

[43] NVIDIA, M. Mittal, P. Roth, J. Tigue, A. Richard, O. Zhang, P. Du, A. Serrano-Muñoz, X. Yao, R. Zurbrügg, N. Rudin, L. Wawrzyniak, M. Rakhsha, A. Denzler, E. Heiden, A. Borovicka, O. Ahmed, I. Akinola, A. Anwar, M.T. Carlson, J.Y. Feng, A. Garg, R. Gasoto, L. Gulich, Y. Guo, M. Gussert, A. Hansen, M. Kulkarni, C. Li, W. Liu, V. Makoviychuk, G. Malczyk, H. Mazhar, M. Moghani, A. Murali, M. Noseworthy, A. Poddubny, N. Ratliff, W. Rehberg, C. Schwarke, R. Singh, J.L. Smith, B. Tang, R. Thaker, M. Trepte, K. Van Wyk, F. Yu, A. Millane, V. Ramasamy, R. Steiner, S. Subramanian, C. Volk, C. Chen, N. Jawale, A.V. Kuruttukulam, M.A. Lin, A. Mandlekar, K. Patzwaldt, J. Welsh, H. Zhao, F. Anes, J.-F. Lafleche, N. Moënne-Loccoz, S. Park, R. Stepinski, D. Van Gelder, C. Amevor, J. Carius, J. Chang, A.H. Chen, P.d. Ciechomski, G. Daviet, M. Mohajerani, J. von Muralt, V. Reutskyy, M. Sauter, S. Schirm, E.L. Shi, P. Terdiman, K. Vilella, T. Widmer, G. Yeoman, T. Chen, S. Grizan, C. Li, L. Li, C. Smith, R. Wiltz, K. Alexis, Y. Chang, D. Chu, L.J. Fan, F. Farshidian, A. Handa, S. Huang, M. Hutter, Y. Narang, S. Pouya, S. Sheng, Y. Zhu, M. Macklin, A. Moravanszky, P. Reist, Y. Guo, D. Hoeller, G. State, Isaac lab: A GPU-accelerated simulation framework for multi-modal robot learning, 2025, http://dx.doi.org/10.48550/ARXIV.2511.04831, Version Number: 1. URL https://arxiv.org/abs/2511.04831.

[44] F. Xiang, Y. Qin, K. Mo, Y. Xia, H. Zhu, F. Liu, M. Liu, H. Jiang, Y. Yuan, H. Wang, L. Yi, A.X. Chang, L.J. Guibas, H. Su, SAPIEN: A SimulAted part-based interactive environment, 2020, http://dx.doi.org/10.48550/ARXIV.2003.08515, Version Number: 1. URL https://arxiv.org/abs/2003.08515.

[45] M. Savva, A. Kadian, O. Maksymets, Y. Zhao, E. Wijmans, B. Jain, J. Straub, J. Liu, V. Koltun, J. Malik, D. Parikh, D. Batra, Habitat: A platform for embodied AI research, in: 2019 IEEE/CVF International Conference on Computer Vision (ICCV), IEEE, Seoul, Korea (South), 2019, pp. 9338–9346, http://dx.doi.org/10.1109/ICCV.2019.00943, URL https://ieeexplore.ieee.org/document/9010745/.

[46] A. Szot, A. Clegg, E. Undersander, E. Wijmans, Y. Zhao, J. Turner, N. Maestre, M. Mukadam, D. Chaplot, O. Maksymets, A. Gokaslan, V. Vondrus, S. Dharur, F. Meier, W. Galuba, A. Chang, Z. Kira, V. Koltun, J. Malik, M. Savva, D. Batra, Habitat 2.0: Training home assistants to rearrange their habitat, in: Proceedings of the 35th International Conference on Neural Information Processing Systems, NIPS '21, Curran Associates Inc., Red Hook, NY, USA, 2021.

[47] X. Puig, E. Undersander, A. Szot, M.D. Cote, T.-Y. Yang, R. Partsey, R. Desai, A.W. Clegg, M. Hlavac, S.Y. Min, V. Vondruš, T. Gervet, V.-P. Berges, J.M. Turner, O. Maksymets, Z. Kira, M. Kalakrishnan, J. Malik, D.S. Chaplot, U. Jain, D. Batra, A. Rai, R. Mottaghi, Habitat 3.0: A co-habitat for humans, avatars and robots, 2023, http://dx.doi.org/10.48550/ARXIV.2310.13724, Version Number: 1. URL https://arxiv.org/abs/2310.13724.

[48] D. Wu, Q. Zhao, Y. Shen, J. Li, P. Zheng, J. Qi, J. Hu, A mixed reality-assisted human-to-robot skill transfer approach for contact-rich assembly via visuomotor primitives, Robot. Comput.-Integr. Manuf. 99 (2026) 103208, http://dx.doi.org/10.1016/j.rcim.2025.103208, URL https://www.sciencedirect.com/science/article/pii/S0736584525002625.

[49] B. Kitchenham, S. Charters, D. Budgen, P. Brereton, M. Turner, S. Linkman, M. Jørgensen, E. Mendes, G. Visaggio, Guidelines for Performing Systematic Literature Reviews in Software Engineering. URL https://legacyfileshare.elsevier.com/promis_misc/525444systematicreviewsguide.pdf.

[50] A. Carrera-Rivera, W. Ochoa, F. Larrinaga, G. Lasa, How-to conduct a systematic literature review: A quick guide for computer science research, MethodsX 9 (2022) 101895, http://dx.doi.org/10.1016/j.mex.2022.101895, URL https://linkinghub.elsevier.com/retrieve/pii/S2215016122002746.

[51] M.J. Page, J.E. McKenzie, P.M. Bossuyt, I. Boutron, T.C. Hoffmann, C.D. Mulrow, L. Shamseer, J.M. Tetzlaff, E.A. Akl, S.E. Brennan, R. Chou, J. Glanville, J.M. Grimshaw, A. Hróbjartsson, M.M. Lalu, T. Li, E.W. Loder, E. Mayo-Wilson, S. McDonald, L.A. McGuinness, L.A. Stewart, J. Thomas, A.C. Tricco, V.A. Welch, P. Whiting, D. Moher, The PRISMA 2020 statement: An updated guideline for reporting systematic reviews, BMJ (2021) n71, http://dx.doi.org/10.1136/bmj.n71.

[52] H. Fan, X. Liu, J.Y.H. Fuh, W.F. Lu, B. Li, Embodied intelligence in manufacturing: Leveraging large language models for autonomous industrial robotics, J. Intell. Manuf. 36 (2) (2025) 1141–1157, http://dx.doi.org/10.1007/s10845-023-02294-y.

[53] C. Margadji, S.W. Pattinson, Hybrid reasoning for perception, explanation, and autonomous action in manufacturing, 2025, http://dx.doi.org/10.48550/ARXIV.2506.08462.

[54] W.-Y. Lin, C.-H. Shih, Empowered smart manufacturing based on an intelligent robotic workcell assistant, in: 2025 International Conference on Advanced Robotics and Intelligent Systems (ARIS), IEEE, Taichung, Taiwan, 2025, pp. 1–7, http://dx.doi.org/10.1109/ARIS66143.2025.11163399.

[55] F. Shi, T. Anzai, Y. Kojio, K. Okada, M. Inaba, Learning agile hybrid whole-body motor skills for thruster-aided humanoid robots, in: 2022 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), IEEE, Kyoto, Japan, 2022, pp. 12986–12993, http://dx.doi.org/10.1109/IROS47612.2022.9981974.

[56] S. Li, Z. Yan, Z. Wang, Y. Gao, VLM-msgraph: Vision language model-enabled multi-hierarchical scene graph for robotic assembly, Robot. Comput.-Integr. Manuf. 94 (2025) 102978, http://dx.doi.org/10.1016/j.rcim.2025.102978.

[57] C. Liu, J. Song, D. Tang, L. Wang, H. Zhu, Q. Cai, From insight to autonomous execution: VLM-enhanced embodied agents towards digital twin-assisted human-robot collaborative assembly, Robot. Comput.-Integr. Manuf. 98 (2026) 103176, http://dx.doi.org/10.1016/j.rcim.2025.103176.

[58] Z. Chen, J. Huo, Y. Chen, Y. Gao, RoboHorizon: An llm-assisted multi-view world model for long-horizon robotic manipulation, 2025, http://dx.doi.org/10.48550/ARXIV.2501.06605.

[59] F. Chen, L. Zou, L. Zou, Research on humanoid robot motion generation and real-time trajectory optimization based on sparse attention mechanism in deep reinforcement learning, Discov. Artif. Intell. 5 (1) (2025) 324, http://dx.doi.org/10.1007/s44163-025-00603-3.

[60] X. Xu, Y. Gu, T. Zhang, J. Yu, S. Skinner, Revolutionizing clean energy labs: Robotic imitation learning for efficient fabrication AI-powered electrical units assembly platform, Energy AI 21 (2025) 100517, http://dx.doi.org/10.1016/j.egyai.2025.100517.

[61] Y. Fan, D. Antonelli, A. Simeone, N. Bao, Intelligent robot assistants for the integration of neurodiverse operators in manufacturing industry, Procedia CIRP 126 (2024) 236–241, http://dx.doi.org/10.1016/j.procir.2024.08.332.

[62] M. Balasubramani, J. Chen, R. Chang, J.-S. Shieh, Development of a human-centric autonomous heating, ventilation, and air conditioning control system enhanced for industry 5.0 chemical fiber manufacturing, Machines 13 (5) (2025) 421, http://dx.doi.org/10.3390/machines13050421.

[63] H. Chu, T. Zhang, Y. Zou, H. Sun, "Follower" to "collaborator": A robot proactive collaborative controller based on human multimodal information for 3D handling/assembly scenarios, J. Manuf. Syst. 73 (2024) 287–306, http://dx.doi.org/10.1016/j.jmsy.2024.02.008.

[64] C. Liu, D. Tang, H. Zhu, Z. Zhang, L. Wang, Y. Zhang, Vision language model-enhanced embodied intelligence for digital twin-assisted human-robot collaborative assembly, J. Ind. Inf. Integr. 48 (2025) 100943, http://dx.doi.org/10.1016/j.jii.2025.100943.

[65] C. Gu, J. Liu, H. Chen, R. Huang, Q. Wuwu, Z. Liu, X. Li, Y. Li, R. Zhang, P. Jia, P.-A. Heng, S. Zhang, Manualvla: A unified VLA model for chain-of-thought manual generation and robotic manipulation, 2025, http://dx.doi.org/10.48550/ARXIV.2512.02013.

[66] S. Li, L. Gao, J. Cao, Y. Hu, Graph-fused vision-language-action for policy reasoning in multi-arm robotic manipulation, 2025, http://dx.doi.org/10.48550/ARXIV.2509.07957.

[67] G. Tang, Q. Jia, G. Chen, Z. Huang, Z. Yao, N. Ji, Three-dimensional-grounded vision-language framework for robotic task planning: Automated prompt synthesis and supervised reasoning, Eng. Appl. Artif. Intell. 164 (2026) 113268, http://dx.doi.org/10.1016/j.engappai.2025.113268.

[68] H. Chen, Y. Yao, R. Liu, C. Liu, J. Ichnowski, Automating robot failure recovery using vision-language models with optimized prompts, 2024, http://dx.doi.org/10.48550/ARXIV.2409.03966.

[69] A. Orsula, M. Geist, M. Olivares-Mendez, C. Martinez, Leveraging procedural generation for learning autonomous peg-in-hole assembly in space, 2024, http://dx.doi.org/10.48550/ARXIV.2405.01134.

[70] B. Chen, G. Xu, L. Wang, C. Jiang, Z. Zhang, Z. Wang, X. Xia, Behavioral planning and parameter meta learning for embodied intelligence robots in adaptive assembly, J. Ind. Inf. Integr. 49 (2026) 100995, http://dx.doi.org/10.1016/j.jii.2025.100995.

[71] B. Kreis, N. Dengler, J.D. Heuvel, R. Menon, H. Perur, M. Bennewitz, Compact multi-object placement using adjacency-aware reinforcement learning, in: 2024 IEEE-RAS 23rd International Conference on Humanoid Robots (Humanoids), IEEE, Nancy, France, 2024, pp. 698–705, http://dx.doi.org/10.1109/Humanoids58906.2024.10769899.

[72] Y. Yuan, S. Wang, Y. Mei, W. Zhang, J. Sun, G. Wang, Improving world models for robot arm grasping with backward dynamics prediction, Int. J. Mach. Learn. Cybern. 15 (9) (2024) 3879–3891, http://dx.doi.org/10.1007/s13042-024-02125-3.

[73] E. Nava, V. Montesinos, E. Bauer, B. Forrai, J. Pai, S. Weirich, S.-D. Gravert, P. Wand, S. Polinski, B.F. Grewe, R.K. Katzschmann, Mimic-one: A scalable model recipe for general purpose robot dexterity, 2025, http://dx.doi.org/10.48550/ARXIV.2506.11916.

[74] L. Tong, K. Song, H. Tian, Y. Man, Y. Yan, Q. Meng, A novel RGB-D cross-background robot grasp detection dataset and background-adaptive grasping network, IEEE Trans. Instrum. Meas. 73 (2024) 1–15, http://dx.doi.org/10.1109/TIM.2024.3413164.

[75] M. Liu, Z. Zhao, W. Zhang, L. Hao, Reinforcement learning control of a humanoid robotic hand actuated by shape memory alloy, Proc. Inst. Mech. Eng. C: J. Mech. Eng. Sci. 235 (21) (2021) 5736–5744, http://dx.doi.org/10.1177/0954406220982019.

[76] Y. Zhang, Y. Liu, S. Liu, W. Liang, C. Wang, K. Wang, Multimodal perception for indoor mobile robotics navigation and safe manipulation, IEEE Trans. Cogn. Dev. Syst. 17 (5) (2025) 1074–1086, http://dx.doi.org/10.1109/TCDS.2024.3481457.

[77] P. Feng, T. Yang, M. Liang, L. Wang, Y. Gao, OC-HMAS: dynamic self-organization and self-correction in heterogeneous multiagent systems using multimodal large models, IEEE Internet Things J. 12 (10) (2025) 13538–13555, http://dx.doi.org/10.1109/JIOT.2025.3545496.

[78] J. Baltes, I. Akbar, S. Saeedvand, Cooperative dual-actor proximal policy optimization algorithm for multi-robot complex control task, Adv. Eng. Informatics 63 (2025) 102960, http://dx.doi.org/10.1016/j.aei.2024.102960.

[79] J. Zhang, S. Schwertfeger, A. Kleiner, From observation to action: Latent action-based primitive segmentation for VLA pre-training in industrial settings, 2025, http://dx.doi.org/10.48550/ARXIV.2511.21428.

[80] Y. Tian, Y. Yang, Y. Xie, Z. Cai, X. Shi, N. Gao, H. Liu, X. Jiang, Z. Qiu, F. Yuan, Y. Li, P. Wang, J. Cai, J. Zeng, H. Dong, J. Pang, InternData-A1: Pioneering high-fidelity synthetic data for pre-training generalist policy, 2025, http://dx.doi.org/10.48550/ARXIV.2511.16651.

[81] J. Lv, Y. Feng, C. Zhang, S. Zhao, L. Shao, C. Lu, SAM-RL: Sensing-aware model-based reinforcement learning via differentiable physics-based simulation and rendering, Int. J. Robot. Res. 44 (10–11) (2025) 1767–1783, http://dx.doi.org/10.1177/02783649241284653.

[82] J. Chen, S. Huang, X. Wang, P. Wang, J. Zhu, Z. Xu, G. Wang, Y. Yan, L. Wang, Perception-decision-execution coordination mechanism driven dynamic autonomous collaboration method for human-like collaborative robot based on multimodal large language model, Robot. Comput.-Integr. Manuf. 98 (2026) 103167, http://dx.doi.org/10.1016/j.rcim.2025.103167.

[83] Y. Xu, R. Bao, L. Zhang, J. Wang, S. Wang, Embodied intelligence in RO/RO logistic terminal: Autonomous intelligent transportation robot architecture, Sci. China Inf. Sci. 68 (5) (2025) 150210, http://dx.doi.org/10.1007/s11432-024-4395-7.

[84] Y. Du, Y. Li, B. Jia, Y. Lin, P. Zhou, W. Liang, Y. Yang, S. Huang, Learning human-humanoid coordination for collaborative object carrying, 2025, http://dx.doi.org/10.48550/ARXIV.2510.14293.

[85] S. Wang, Z. Tan, Y. Wang, Z. Zhou, D. Zhu, A novel paradigm of robotic machining towards embodied intelligent manufacturing: case study on paint defect repair, Robot. Comput.-Integr. Manuf. 98 (2026) 103184, http://dx.doi.org/10.1016/j.rcim.2025.103184.

[86] A. Tafuro, L. Cacciani, A.M. Zanchettin, P. Rocco, Towards intelligent robotic sole deburring: From burrs identification to path planning, in: 2024 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), IEEE, Abu Dhabi, United Arab Emirates, 2024, pp. 7189–7195, http://dx.doi.org/10.1109/IROS58592.2024.10802795.

[87] J. Sousa, B. Brandau, R. Hemschik, R. Darabi, A. Sousa, L.P. Reis, F. Brueckner, A. Reis, JEMA-SINDYc: End-to-end control using joint embedding multimodal alignment in directed energy deposition, Addit. Manuf. 109 (2025) 104888, http://dx.doi.org/10.1016/j.addma.2025.104888.

[88] H. Huang, C.E. Mower, G. Huang, S. Das, M. Dierking, G. Luo, K. Tan, X. Chen, Y. Yang, Y. Chen, Y. Zeng, Y. Li, Z. Zhang, S. Wu, Y. Zhang, W. Qiu, T. Cao, M. Qin, S. Pakdamansavoji, Y. Liu, Y. Zhuang, G. Tian, J. Hao, J. Wang, H. Bou-Ammar, X. Quan, OpenPyRo-A1: An open python-based low-cost bimanual robot for embodied AI, IEEE Robot. Autom. Lett. 11 (1) (2026) 746–753, http://dx.doi.org/10.1109/LRA.2025.3634886.

[89] N. Adiuku, N.P. Avdelidis, G. Tang, A. Plastropoulos, Y. Diallo, Mobile robot obstacle detection and avoidance with NAV-YOLO, Int. J. Mech. Eng. Robot. Res. 13 (2) (2024) 219–226, http://dx.doi.org/10.18178/ijmerr.13.2.219-226.

[90] S. Li, Z. Wang, Z. Yan, Y. Gao, H. Jiang, P. Zheng, Large language model for humanoid cognition in proactive human-robot collaboration, in: 2024 IEEE 20th International Conference on Automation Science and Engineering (CASE), IEEE, Bari, Italy, 2024, pp. 540–545, http://dx.doi.org/10.1109/CASE59546.2024.10711379.

[91] Y. Peng, Z. Wang, Y. Zhang, S. Zhang, N. Cai, F. Wu, M. Chen, Revolutionizing battery disassembly: The design and implementation of a battery disassembly autonomous mobile manipulator robot(BEAM-1), 2024, http://dx.doi.org/10.48550/ARXIV.2407.06590.

[92] M. Trovinger, L. Giacobbe, T. Grantham, C. Sun, J. Chen, Reinforcement learning with vision-based state estimation for generalized robotic assembly, in: 42nd International Symposium on Automation and Robotics in Construction, Montreal, Canada, 2025, http://dx.doi.org/10.22260/ISARC2025/0025.

[93] International Organization for Standardization, ISO 10218-1:2025 Robotics — Safety requirements — Part 1: Industrial Robots, Standard, 2025.

[94] International Organization for Standardization, ISO 10218-2:2025 Robotics — Safety requirements — Part 2: Industrial robot applications and robot cells, Standard, 2025.

[95] International Organization for Standardization, ISO/TS 15066:2016 Robots and Robotic Devices — Collaborative Robots, Technical Specification, 2016.

[96] International Organization for Standardization, ISO 13849-1:2023 Safety of Machinery — Safety-Related Parts of Control Systems — Part 1: General Principles for Design, Standard, 2023.

[97] International Electrotechnical Commission, Standard, 2021, URL https://webstore.iec.ch/publication/59927.

[98] European Parliament and Council of the European Union, Regulation (EU) 2023/1230 of the European parliament and of the council of 14 June 2023 on machinery and repealing directive 2006/42/EC, 2023, Official Journal of the European Union, L 165, pp. 1–102. URL https://eur-lex.europa.eu/eli/reg/2023/1230/oj.

[99] European Parliament and Council of the European Union, Regulation (EU) 2024/1689 of the European parliament and of the council of 13 June 2024 laying down harmonised rules on artificial intelligence (artificial intelligence act), 2024, Official Journal of the European Union, L, 2024/1689. URL https://eur-lex.europa.eu/eli/reg/2024/1689/oj.

[100] P. Chi, Z. Wang, H. Liao, T. Li, Q. Zhang, A 3d reconstruction and relocalization method for humanoid welding robots, IEEE Signal Process. Lett. 32 (2025) 1071–1075, http://dx.doi.org/10.1109/LSP.2025.3544967.

[101] C. Bao, T.-H. Kim, A. Hassanpoor Kalhori, W.S. Kim, A 3D-printed neuromorphic humanoid hand for grasping unknown objects, IScience 25 (4) (2022) 104119, http://dx.doi.org/10.1016/j.isci.2022.104119.

[102] Z. Chen, S. Zhong, J. Bai, C. Zhang, H. Qiao, A force-position hybrid control for full-size humanoid robots based on adaptive parameters, IEEE/ASME Trans. Mechatronics (2025) 1–11, http://dx.doi.org/10.1109/TMECH.2025.3583018.

[103] G. Nadizar, E. Medvet, K. Walker, S. Risi, A fully-distributed shape-aware neural controller for modular robots, in: Proceedings of the Genetic and Evolutionary Computation Conference, ACM, Lisbon Portugal, 2023, pp. 184–192, http://dx.doi.org/10.1145/3583131.3590419.

[104] L. Zhang, H. Zhan, X. Liu, H. Cao, F. Xing, Z. You, A planar compound eye based microsystem for high precision 3D perception, PhotoniX 5 (1) (2024) 21, http://dx.doi.org/10.1186/s43074-024-00136-5.

[105] H. Tang, Y. Li, N. Zhou, M. Cheng, S. Tao, B. Xu, A real-time perception–motion codesign method for image-based visual servoing in embodied intelligence systems, J. Ind. Inf. Integr. 48 (2025) 100933, http://dx.doi.org/10.1016/j.jii.2025.100933.

[106] O. Jorg, G. Fantoni, A self-adaptive high precision gripper for shape variant components: towards higher reliability and efficiency of a cobotic cell, J. Manuf. Syst. 70 (2023) 113–126, http://dx.doi.org/10.1016/j.jmsy.2023.04.003.

[107] R.J. Nash, Y. Yang, Y. Li, Adaptive mechanical metamaterials with on-demand binary local modulus for embodied intelligence, Adv. Sci. 12 (40) (2025) e09675, http://dx.doi.org/10.1002/advs.202509675.

[108] Y. Xian, K. Huang, Z. Zhu, S. Zhen, Y.-H. Chen, Adaptive robust control for fuzzy mechanical systems in confined spaces: Nash game optimization design, IEEE Trans. Fuzzy Syst. 32 (5) (2024) 2863–2875, http://dx.doi.org/10.1109/TFUZZ.2024.3363901.

[109] Y. Liu, H. Wang, D. Wang, Y. Li, X. Zhang, Advanced structural design for humanoid robot upper limbs, in: 2025 9th International Conference on Robotics, Control and Automation (ICRCA), IEEE, Shanghai, China, 2025, pp. 33–38, http://dx.doi.org/10.1109/ICRCA64997.2025.11011241.

[110] M.L. Romero, R. Suyama, Agentic AI for intent-based industrial automation, in: 2025 16th IEEE International Conference on Industry Applications (INDUSCON), IEEE, São Sebastião, Brazil, 2025, pp. 437–444, http://dx.doi.org/10.1109/INDUSCON66435.2025.11241317.

[111] J. Lee, H. Su, Agentic AI for smart manufacturing, Manuf. Lett. 46 (2025) 92–96, http://dx.doi.org/10.1016/j.mfglet.2025.10.013.

[112] A. Fernández-Miguel, S. Ortíz-Marcos, M. Jiménez-Calzado, A.P. Fernández Del Hoyo, F.E. García-Muiña, D. Settembre-Blundo, Agentic AI in smart manufacturing: Enabling human-centric predictive maintenance ecosystems, Appl. Sci. 15 (21) (2025) 11414, http://dx.doi.org/10.3390/app152111414.

[113] P. Chi, Z. Wang, H. Liao, T. Li, X. Wu, J. Wu, Q. Zhang, An integrated calibration and 3D reconstruction method for humanoid welding robots, Measurement 253 (2025) 117748, http://dx.doi.org/10.1016/j.measurement.2025.117748.

[114] F. Gosselin, G. Acher, B. Gradoussoff, S. Kchir, F. Keith, O. Lebec, C. Louison, B. Luvison, F.M. De Chamisso, B. Meden, V. Molina, M. Morelli, J. Rabarisoa, C. Vienne, G. Ameyugo, An intelligent robotics modular architecture for easy adaptation to novel tasks and applications, in: 2023 IEEE 19th International Conference on Automation Science and Engineering (CASE), IEEE, Auckland, New Zealand, 2023, pp. 1–8, http://dx.doi.org/10.1109/CASE56687.2023.10260553.

[115] O. Herzog, S. Nertinger, K. Wenzel, A. Naceri, S. Haddadin, K. Bengler, Assessing perceived discomfort and proxemic behavior towards robots: A comparative study between real and augmented reality presentations, in: 2023 32nd IEEE International Conference on Robot and Human Interactive Communication (RO-MAN), IEEE, Busan, Korea, Republic of, 2023, pp. 959–965, http://dx.doi.org/10.1109/RO-MAN57019.2023.10309413.

[116] Z. Sun, H. Yang, Q. Dong, Y. Mo, H. Li, Z. Jiang, Autonomous assembly method of 3-arm robot to fix the multipin and hole load plate on a space station, Space: Sci. Technol. 2021 (2021) http://dx.doi.org/10.34133/2021/9815389, 2021/9815389.

[117] H. Wang, Y. Xu, Z. Xu, J. Gao, Y. Liu, W. Hu, K. Wang, Z. Zhang, AutoPrune: Each complexity deserves a pruning policy, 2025, http://dx.doi.org/10.48550/ARXIV.2509.23931.

[118] X. Li, G. Hou, Y. Song, Y. Li, Balance control of a wheeled humanoid robot during the standing-up process using backpropagation neural networks, in: 2025 5th International Conference on Neural Networks, Information and Communication Engineering (NNICE), IEEE, Guangzhou, China, 2025, pp. 1345–1349, http://dx.doi.org/10.1109/NNICE64954.2025.11064689.

[119] R. Carlet, K. Gilday, J. Hughes, Behaviour range optimization of the dexterous robotic open parametric hand, in: 2025 IEEE 8th International Conference on Soft Robotics (RoboSoft), IEEE, Lausanne, Switzerland, 2025, pp. 1–7, http://dx.doi.org/10.1109/RoboSoft63089.2025.11020976.

[120] H. Wang, Y. Wang, X. Li, Y. Li, S. Wang, BookBot: A robotic manipulation benchmark for voice-driven book recognition and grasping in cluttered environments, in: 2025 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), IEEE, Hangzhou, China, 2025, pp. 20960–20967, http://dx.doi.org/10.1109/IROS60139.2025.11246718.

[121] T. Li, D. Li, C. Zhang, P. Chi, Z. Luo, Category-controllable and high-fidelity 3D defect synthesis for embodied intelligence-based industrial inspection, J. Ind. Inf. Integr. 49 (2026) 101016, http://dx.doi.org/10.1016/j.jii.2025.101016.

[122] H. Ma, Q. Jiang, X. Ma, R. Chen, K. Hua, X. Yang, J. Ge, J. Ji, M. Xue, Coaxial graphene/mxene microfibers with interfacial buffer-based lightweight distance sensors assisting lossless grasping of fragile and deformable objects, Langmuir 39 (12) (2023) 4530–4536, http://dx.doi.org/10.1021/acs.langmuir.3c00374.

[123] V. Kekshin, S. Ponomar, M. Sarkans, V. Kuts, S. Pavlov, Collaboration between industrial, collaborative, humanoid robots and humans, J. Mach. Eng. 25 (2) (2025) 100–110, http://dx.doi.org/10.36897/jme/203790.

[124] G. Zambella, G. Caporale, G. Grioli, L. Pallottino, A. Bicchi, Composite whole-body control of two-wheeled robots, IEEE Trans. Robot. 41 (2025) 2301–2321, http://dx.doi.org/10.1109/TRO.2025.3548494.

[125] F. Wang, S. Liu, Conceptual framework toward embodied collective adaptive intelligence, 2025, http://dx.doi.org/10.48550/ARXIV.2505.23153.

[126] W. Li, X. Zhang, Cooperative motion planning for mobile humanoid robots based on the manipulability map, in: 2024 IEEE International Conference on Mechatronics and Automation (ICMA), IEEE, Tianjin, China, 2024, pp. 685–690, http://dx.doi.org/10.1109/ICMA61710.2024.10632896.

[127] J. Chai, Y. Li, E. Dong, Customize-your-joy hand: A user-oriented, cost-effective 22-DOF platform for future human-robot community, IEEE Robot. Autom. Lett. 10 (12) (2025) 13344–13351, http://dx.doi.org/10.1109/LRA.2025.3612226.

[128] Y. Li, D. Zheng, L. Jiang, Y. Liu, Y. Du, M. Cong, Design and evaluation of a rigid-soft coupling anthropomorphic hand system for grasping performance enhancement, IEEE Trans. Ind. Inform. 21 (7) (2025) 5755–5765, http://dx.doi.org/10.1109/TII.2025.3558335.

[129] E. Ottonello, M. Baggetta, G. Berselli, A. Parmiggiani, Design and validation of a push-latch gripper made in additive manufacturing, IEEE/ASME Trans. Mechatronics 28 (4) (2023) 2083–2091, http://dx.doi.org/10.1109/TMECH.2023.3276073.

[130] D. Li, Q. Chen, Y. Gao, J. Zhang, W. Hao, Design and validation of a tendon-driven hybrid rigid-soft three-finger gripper for humanoid robots, in: 2025 IEEE 20th Conference on Industrial Electronics and Applications (ICIEA), IEEE, Yantai, China, 2025, pp. 1–6, http://dx.doi.org/10.1109/ICIEA65512.2025.11149233.

[131] F. Stroppa, Design optimizer for planar soft-growing robot manipulators, Eng. Appl. Artif. Intell. 130 (2024) 107693, http://dx.doi.org/10.1016/j.engappai.2023.107693.

[132] T. Xiang, B. Li, X. Pan, Q. Zhang, Development of a simple and novel digital twin framework for industrial robots in intelligent robotics manufacturing, in: 2024 IEEE 20th International Conference on Automation Science and Engineering (CASE), IEEE, Bari, Italy, 2024, pp. 4187–4193, http://dx.doi.org/10.1109/CASE59546.2024.10711459.

[133] W. Chen, C. Xiao, G. Gao, F. Sun, C. Zhang, J. Zhang, DreamArrangement: Learning language-conditioned robotic rearrangement of objects via denoising diffusion and vlm planner, IEEE Trans. Syst. Man, Cybern.: Syst. 55 (11) (2025) 8675–8688, http://dx.doi.org/10.1109/TSMC.2025.3611698.

[134] Y. Qin, A. Escande, F. Kanehiro, E. Yoshida, Dual-arm mobile manipulation planning of a long deformable object in industrial installation, IEEE Robot. Autom. Lett. 8 (5) (2023) 3039–3046, http://dx.doi.org/10.1109/LRA.2023.3264779.

[135] Y. Tang, L. He, Dual-arm trajectory tracking algorithm for humanoid robot based on fully-actuated system approach, J. Mech. Sci. Technol. 39 (10) (2025) 6205–6219, http://dx.doi.org/10.1007/s12206-025-0536-3.

[136] S.M.M. Rahman, Dynamic affect-based motion planning of a humanoid robot for human-robot collaborative assembly in manufacturing, Electronics 13 (6) (2024) 1044, http://dx.doi.org/10.3390/electronics13061044.

[137] Q. Li, F. Meng, Z. Yu, X. Chen, Q. Huang, Dynamic torso compliance control for standing and walking balance of position-controlled humanoid robots, IEEE/ASME Trans. Mechatronics 26 (2) (2021) 679–688, http://dx.doi.org/10.1109/TMECH.2021.3061825.

[138] A. Bazhenov, S. Satsevich, S. Egorov, F. Khabibullin, D. Tsetserukou, Echo: An open-source, low-cost teleoperation system with force feedback for dataset collection in robot learning, 2025, http://dx.doi.org/10.48550/ARXIV.2504.07939.

[139] M. Pantano, A. Curioni, D. Regulin, T. Kamps, D. Lee, Effects of robotic expertise and task knowledge on physical ergonomics and joint efficiency in a human-robot collaboration task, in: 2023 IEEE-RAS 22nd International Conference on Humanoid Robots (Humanoids), IEEE, Austin, TX, USA, 2023, pp. 1–8, http://dx.doi.org/10.1109/Humanoids57100.2023.10375163.

[140] Y. Gan, B. Zhang, J. Shao, Z. Han, A. Li, X. Dai, Embodied intelligence: Bionic robot controller integrating environment perception, autonomous planning, and motion control, IEEE Robot. Autom. Lett. 9 (5) (2024) 4559–4566, http://dx.doi.org/10.1109/LRA.2024.3377559.

[141] T. Yue, C. Lu, K. Tang, Q. Qi, Z. Lu, L.Y. Lee, H. Bloomfield-Gadelha, J. Rossiter, Embodying soft robots with octopus-inspired hierarchical suction intelligence, Sci. Robot. 10 (102) (2025) eadr4264, http://dx.doi.org/10.1126/scirobotics.adr4264.

[142] A.C. Bavelos, N. Kousi, C. Gkournelos, K. Lotsaris, S. Aivaliotis, G. Michalos, S. Makris, Enabling flexibility in manufacturing by integrating shopfloor and process perception for mobile robot workers, Appl. Sci. 11 (9) (2021) 3985, http://dx.doi.org/10.3390/app11093985.

[143] S. Ganatra, H. Jeong, Enhancing human-robot teleoperation through depth sensing and extended reality, in: 2025 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW), IEEE, Saint Malo, France, 2025, pp. 1110–1114, http://dx.doi.org/10.1109/VRW66409.2025.00221.

[144] S. Mukherjee, S. Aine, M. Likhachev, ePA*SE: Edge-based parallel A* for slow evaluations, 2022, http://dx.doi.org/10.48550/ARXIV.2203.01369.

[145] S. Mannem, W. Macke, P. Stone, R. Mirsky, Exploring the cost of interruptions in human-robot teaming, in: 2023 IEEE-RAS 22nd International Conference on Humanoid Robots (Humanoids), IEEE, Austin, TX, USA, 2023, pp. 1–8, http://dx.doi.org/10.1109/Humanoids57100.2023.10375236.

[146] H. Park, Y. Jeong, W. Kim, J. Choi, J. Ahn, J.-H. Jeong, I. Park, J. Kim, Field-programmable robotic folding sheet, Nat. Commun. 16 (1) (2025) 6937, http://dx.doi.org/10.1038/s41467-025-61838-3.

[147] A. Moteki, S. Masui, F. Yang, Y. Song, Y. Bisk, G. Neubig, I. Kusajima, Y. Watanabe, H. Ishida, J. Takahashi, S. Jiang, FieldWorkArena: Agentic AI benchmark for real field work tasks, 2025, http://dx.doi.org/10.48550/ARXIV.2505.19662.

[148] C. Lee, H. Kim, M. Oh, K. Ok, S.-H. Ahn, Frictional and prismatic pin-array gripper for universal gripping and stable tool manipulation, IEEE Trans. Robot. 41 (2025) 6693–6706, http://dx.doi.org/10.1109/TRO.2025.3626656.

[149] S. Dissanayaka, A. Ferrein, T. Hofmann, K. Nakajima, M. Sanz-Lopez, J. Savage, D. Swoboda, M. Tschesche, W. Uemura, T. Viehmann, S. Yasuda, From production logistics to smart manufacturing: The vision for a new RoboCup industrial league, 2025, http://dx.doi.org/10.48550/ARXIV.2507.11402.

[150] H. Li, H. Madhukumar, N. Methley, X. Chen, Y. Wu, J. Parra-Ullauri, V. Sharma, J. Lee, A.R. Koblitz, M. Andrews, S. Liu, Y. Deng, O.Y. Kolawole, A. Tassi, D. Simeonidou, Future factories with 6G: Agentic AI and cyber-physical digital twins, IEEE Internet Things J. (2025) http://dx.doi.org/10.1109/JIOT.2025.3623075, 1–1.

[151] Y. Dong, X. Liu, J. Wan, Z. Deng, GEX: Democratizing dexterity with fully-actuated dexterous hand and exoskeleton glove, 2025, http://dx.doi.org/10.48550/ARXIV.2506.04982.

[152] X. Xie, Y. Qin, Z. Zhang, Z. Yan, H. Jin, M. Xu, C. Zhang, GY-SLAM: A dense semantic SLAM system for plant factory transport robots, Sensors 24 (5) (2024) 1374, http://dx.doi.org/10.3390/s24051374.

[153] H. Yang, X. Meng, J. Liang, Y. Zhang, K. Li, HCDA: A hidden cross-domain authentication protocol for embodied intelligence in smart manufacturing, J. Ind. Inf. Integr. 48 (2025) 100946, http://dx.doi.org/10.1016/j.jii.2025.100946.

[154] M.S. Bajestani, M.M. Mahdi, D. Mun, D.B. Kim, Human and humanoid-in-the-loop (hhitl) ecosystem: An industry 5.0 perspective, Machines 13 (6) (2025) 510, http://dx.doi.org/10.3390/machines13060510.

[155] X. Lin, Y. Zhang, Y. Gan, J. Chen, H. Shen, Y. He, L. Li, Z. Yuan, S. Wang, C. Wang, R. Zhang, N. Li, J. Liu, Human-AI co-embodied intelligence for scientific experimentation and manufacturing, 2025, http://dx.doi.org/10.48550/ARXIV.2511.02071.

[156] D. Hostettler, S. Mayer, C. Hildebrand, Human-like movements of industrial robots positively impact observer perception, Int. J. Soc. Robot. 15 (8) (2023) 1399–1417, http://dx.doi.org/10.1007/s12369-022-00954-2.

[157] C. Dai, C. Ye, J. Ren, S. Yang, L. Cao, H. Yu, S. Liu, Z. Shao, J. Li, W. Chen, S. Ling, Humanoid ionotronic skin for smart object recognition and sorting, ACS Mater. Lett. 5 (1) (2023) 189–201, http://dx.doi.org/10.1021/acsmaterialslett.2c00783.

[158] K. Chappellet, M. Murooka, G. Caron, F. Kanehiro, A. Kheddar, Humanoid loco-manipulations using combined fast dense 3D tracking and SLAM with wide-angle depth-images, IEEE Trans. Autom. Sci. Eng. 21 (3) (2024) 3691–3704, http://dx.doi.org/10.1109/TASE.2023.3283497.

[159] M.A. Farahani, M.I. Khan, T. Wuest, Hybrid agentic AI and multi-agent systems in smart manufacturing, 2025, http://dx.doi.org/10.48550/ARXIV.2511.18258.

[160] R. Younes, F. Elisei, D. Pellier, G. Bailly, Impact of verbal instructions and deictic gestures of a cobot on the performance of human coworkers, in: 2024 IEEE-RAS 23rd International Conference on Humanoid Robots (Humanoids), IEEE, Nancy, France, 2024, pp. 1040–1047, http://dx.doi.org/10.1109/Humanoids58906.2024.10769925.

[161] D. Crivelli, M. Balconi, Industrial collaborative robotics, humanoid cobots, and social embodiment: A critical look on human–robot interaction, Int. J. Humanoid Robot. 22 (01) (2025) 2430001, http://dx.doi.org/10.1142/S0219843624300010.

[162] L. Ren, H. Wang, J. Dong, Z. Jia, S. Li, Y. Wang, Y. Laili, D. Huang, L. Zhang, B. Li, Industrial foundation model, IEEE Trans. Cybern. 55 (5) (2025) 2286–2301, http://dx.doi.org/10.1109/TCYB.2025.3527632.

[163] M.R. Rahul, S.S. Chiddarwar, Integrating virtual twin and deep neural networks for efficient and energy-aware robotic deburring in industry 4.0, Int. J. Precis. Eng. Manuf. 24 (9) (2023) 1517–1534, http://dx.doi.org/10.1007/s12541-023-00875-8.

[164] Z.-Y. Deng, L.-W. Kang, H.-H. Chiang, H.-C. Li, Integration of robotic vision and automatic tool changer based on sequential motion primitive for performing assembly tasks, IFAC-PapersOnLine 56 (2) (2023) 5320–5325, http://dx.doi.org/10.1016/j.ifacol.2023.10.175.

[165] A.A. Malik, T. Masood, A. Brem, Intelligent humanoids in manufacturing to address worker shortage and skill gaps: Case of tesla optimus, 2023, http://dx.doi.org/10.48550/ARXIV.2304.04949.

[166] K. Seepanomwan, D. Caligiore, K.J. O'Regan, G. Baldassarre, Intrinsic motivations and planning to explain tool-use development: A study with a simulated robot model, IEEE Trans. Cogn. Dev. Syst. 14 (1) (2022) 75–89, http://dx.doi.org/10.1109/TCDS.2020.2986411.

[167] B.A. Richardson, F. Grüninger, L. Mack, J. Stueckler, K.J. Kuchenbecker, Isyhand: A dexterous multi-finger robot hand with an articulated palm, in: 2025 IEEE-RAS 24th International Conference on Humanoid Robots (Humanoids), IEEE, Seoul, Korea, Republic of, 2025, pp. 720–727, http://dx.doi.org/10.1109/Humanoids65713.2025.11203015.

[168] Z. Zhou, H. Sun, X.V. Wang, Z. Zhang, Q. Cao, Learning accurate and efficient three-finger grasp generation in clutters with an auto-annotated large-scale dataset, Robot. Comput.-Integr. Manuf. 91 (2025) 102822, http://dx.doi.org/10.1016/j.rcim.2024.102822.

[169] Y. Wu, L. Chen, I.P. Mahiques, Z. Bing, F. Wu, A. Knoll, S. Haddadin, Learning dynamic robot-to-robot object handover, IFAC-PapersOnLine 56 (2) (2023) 4368–4374, http://dx.doi.org/10.1016/j.ifacol.2023.10.1814.

[170] W. Fan, H. Li, Q. Cong, D. Zhang, MagicGripper: A mini-MagicTac integrated gripper enabling multimodal perception in contact-rich manipulation, IEEE Trans. Autom. Sci. Eng. 22 (2025) 24311–24332, http://dx.doi.org/10.1109/TASE.2025.3631485.

[171] W. Fan, H. Li, D. Zhang, MagicGripper: A multimodal sensor-integrated gripper for contact-rich robotic manipulation, 2025, http://dx.doi.org/10.48550/ARXIV.2505.24382.

[172] Z. Yang, Y. Shen, Y. Li, M. Li, L. Sun, MAPF-world: Action world model for multi-agent path finding, 2025, http://dx.doi.org/10.48550/ARXIV.2508.12087.

[173] Q. Xu, R. Lai, J. Yin, Markerless force estimation via SuperPoint-SIFT fusion and finite element analysis: A sensorless solution for deformable object manipulation, Biomimetics 10 (9) (2025) 600, http://dx.doi.org/10.3390/biomimetics10090600.

[174] J. Jiang, F.J. Xi, J. You, Q. Xue, Mechanical design of a new anthropomorphic robot for fastening in wing-box, in: Volume 8B: 45th Mechanisms and Robotics Conference (MR), American Society of Mechanical Engineers, Virtual, Online, 2021, V08BT08A016, http://dx.doi.org/10.1115/DETC2021-68098.

[175] L. Zong, C. Jiang, S. Zhen, C. Ma, Model reconstruction-based optimal adaptive prescribed time control for multi-stage precision robotic arms: Inequality constraints, uncertainties, disturbances, IEEE Trans. Autom. Sci. Eng. 22 (2025) 17147–17159, http://dx.doi.org/10.1109/TASE.2025.3582955.

[176] S. Mukherjee, S. Aine, M. Likhachev, MPLP: Massively parallelized lazy planning, 2021, http://dx.doi.org/10.48550/ARXIV.2107.02826.

[177] P.K. Said, S.R. Mali, N.R. Shilotri, Multi-modal sensing and perception for humanoid service robots, in: 2023 10th IEEE Uttar Pradesh Section International Conference on Electrical, Electronics and Computer Engineering (UPCON), IEEE, Gautam Buddha Nagar, India, 2023, pp. 1217–1221, http://dx.doi.org/10.1109/UPCON59197.2023.10434302.

[178] Z. Xu, Y. Yan, X. Wang, Y. Chen, Q. Xu, Natural entanglement inspired cilia-like soft gripper for rapid adaptive grasping, Adv. Intell. Syst. (2025) 2500468, http://dx.doi.org/10.1002/aisy.202500468.

[179] D. Hanson, A. Imran, G. Morales, V. Krisciunas, A. Sagi, A. Malali, R. Mohbe, R. Upadrashta, Open arms: Open-source arms, hands & control, 2022, http://dx.doi.org/10.48550/ARXIV.2205.12992.

[180] P. Pikulin, V. Lishunov, K. Kułakowski, Optimizing path planning for automated guided vehicles in constrained warehouse environments: Addressing the challenges of non-rotary platforms and irregular layouts, Robotics 14 (4) (2025) 39, http://dx.doi.org/10.3390/robotics14040039.

[181] T. Godden, B.W. Mulvey, E. Redgrave, T. Nanayakkara, Pats-wheel: A passively-transformable single-part wheel for mobile robot navigation on unstructured terrain, IEEE Robot. Autom. Lett. 9 (6) (2024) 5512–5519, http://dx.doi.org/10.1109/LRA.2024.3389828.

[182] X. Tu, L. Fang, H. Zhang, Z. Wang, C. Chen, L. Wang, W. He, H. Liu, P. Wang, Performance-enhanced flexible self-powered tactile sensor arrays based on lotus root-derived porous carbon for real-time human–machine interaction of the robotic snake, ACS Appl. Mater. & Interfaces 16 (7) (2024) 9333–9342, http://dx.doi.org/10.1021/acsami.3c18714.

[183] F. Bousetouane, Physical AI agents: Integrating cognitive intelligence with real-world action, 2025, http://dx.doi.org/10.48550/ARXIV.2501.08944.

[184] W. Wan, K. Harada, F. Kanehiro, Preparatory manipulation planning using automatically determined single and dual arm, IEEE Trans. Ind. Inform. 16 (1) (2020) 442–453, http://dx.doi.org/10.1109/TII.2019.2892772.

[185] I. Matsuo, T. Shimizu, Y. Nakai, M. Kakimoto, Y. Sawasaki, Y. Mori, T. Sugano, S. Ikemoto, T. Miyamoto, Q-Bot: Heavy object carriage robot for in-house logistics based on universal vacuum gripper, Adv. Robot. 34 (3–4) (2020) 173–188, http://dx.doi.org/10.1080/01691864.2019.1709987.

[186] X. Yang, W. Zeng, Y. Zhang, R. Zhu, Z. Ren, Z. Yu, Research on human-computer interaction interface design based on user experience, in: 2024 International Conference on Artificial Intelligence and Power Systems (AIPS), IEEE, Chengdu, China, 2024, pp. 73–77, http://dx.doi.org/10.1109/AIPS64124.2024.00023.

[187] K. Zhao, Research on human-machine collaboration system for intelligent manufacturing - based on multi-source data fusion, in: Proceedings of the 2nd Guangdong-Hong Kong-Macao Greater Bay Area International Conference on Digital Economy and Artificial Intelligence, ACM, Dongguan China, 2025, pp. 1409–1413, http://dx.doi.org/10.1145/3745238.3745458.

[188] C. Zhai, Z. Gong, Y. Li, X. Ding, Y. Huang, Y. Zhang, Research on the application of embodied intelligent robots in logistics support, in: 2025 5th International Conference on Computer, Control and Robotics (ICCCR), IEEE, Hangzhou, China, 2025, pp. 110–115, http://dx.doi.org/10.1109/ICCCR65461.2025.11072625.

[189] Y. Yuan, K. Kitani, Residual force control for agile human behavior imitation and extended motion synthesis, 2020, http://dx.doi.org/10.48550/ARXIV.2006.07364.

[190] I. Lanese, U. Schultz, I. Ulidowski, Reversible execution for robustness in embodied AI and industrial robots, IT Prof. 23 (3) (2021) 12–17, http://dx.doi.org/10.1109/MITP.2021.3073757.

[191] S.H. Bang, C.A. Jové, L. Sentis, RL-augmented MPC framework for agile and robust bipedal footstep locomotion planning and control, in: 2024 IEEE-RAS 23rd International Conference on Humanoid Robots (Humanoids), IEEE, Nancy, France, 2024, pp. 607–614, http://dx.doi.org/10.1109/Humanoids58906.2024.10769914.

[192] H. Tan, X. Hao, C. Chi, M. Lin, Y. Lyu, M. Cao, D. Liang, Z. Chen, M. Lyu, C. Peng, C. He, Y. Ao, Y. Lin, P. Wang, Z. Wang, S. Zhang, Roboos: A hierarchical embodied framework for cross-embodiment and multi-agent collaboration, 2025, http://dx.doi.org/10.48550/ARXIV.2505.03673.

[193] Y. Li, Z. Xu, S. Li, Z. Liao, S. Li, X. Zhou, Robot compliance control framework for grinding thin-walled parts with unknown surface: Deformation and orientation adaptation, Robot. Comput.-Integr. Manuf. 98 (2026) 103147, http://dx.doi.org/10.1016/j.rcim.2025.103147.

[194] A. Ajith, G. Narayanan, J. Zornow, C. Calle, A.H. Lugo, J.L.S. Rincon, C. Wen, E. Solowjow, Robotic automation in apparel manufacturing: A novel approach to fabric handling and sewing, in: 2025 IEEE 21st International Conference on Automation Science and Engineering (CASE), IEEE, Los Angeles, CA, USA, 2025, pp. 3254–3260, http://dx.doi.org/10.1109/CASE58245.2025.11163876.

[195] M. Kim, H. Park, J. Kim, H. Chang, Y. Na, Scalable inflatable origami robot modules with proprioceptive closed-loop shape control, IEEE Robot. Autom. Lett. 10 (10) (2025) 10346–10353, http://dx.doi.org/10.1109/LRA.2025.3601070.

[196] A. Wiedholz, S. Wucherer, S. Dietrich, F. Kerber, Semantic 3D scene segmentation for robotic assembly process execution, in: 2023 IEEE 19th International Conference on Automation Science and Engineering (CASE), IEEE, Auckland, New Zealand, 2023, pp. 1–6, http://dx.doi.org/10.1109/CASE56687.2023.10260532.

[197] S. Duan, F. De Pace, F.P. Sanches, H. Jiang, M. Liarokapis, Semi-autonomous, virtual reality based robotic telemanipulation for the execution of peg-in-hole assembly tasks, in: 2024 IEEE-RAS 23rd International Conference on Humanoid Robots (Humanoids), IEEE, Nancy, France, 2024, pp. 351–358, http://dx.doi.org/10.1109/Humanoids58906.2024.10769787.

[198] P.S. Matharu, A.P. Singh, Y. Song, U. Gandhi, Y. Tadesse, Single layered soft skin actuated with twisted and mandrel coiled actuators for soft robotics, J. Polym. Sci. 62 (10) (2024) 2071–2093, http://dx.doi.org/10.1002/pol.20230691.

[199] E. Hudson, G. Warnell, F. Torabi, P. Stone, Skeletal feature compensation for imitation learning with embodiment mismatch, in: 2022 International Conference on Robotics and Automation (ICRA), IEEE, Philadelphia, PA, USA, 2022, pp. 2482–2488, http://dx.doi.org/10.1109/ICRA46639.2022.9812127.

[200] M. Majewski, I. Czarnowski, Smart factory humanoid robots with cognitive architectures for human-inspired tasks, in: IEEE EUROCON 2025 - 21st International Conference on Smart Technologies, IEEE, Gdynia, Poland, 2025, pp. 1–6, http://dx.doi.org/10.1109/EUROCON64445.2025.11073496.

[201] C. Sirithunge, V. Chandiramani, Y. Xie, H. Hauser, A. Conn, F. Iida, Soft-rigid coupled blade leg achieves spatio-temporal terrain classification with minimal sensor configuration, in: 2025 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), IEEE, Hangzhou, China, 2025, pp. 6016–6023, http://dx.doi.org/10.1109/IROS60139.2025.11247489.

[202] C. Perrot, O. Stasse, Step toward deploying the torque-controlled robot TALOS on industrial operations, in: 2023 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), IEEE, Detroit, MI, USA, 2023, pp. 10405–10411, http://dx.doi.org/10.1109/IROS55552.2023.10342428.

[203] Y. Deng, T. Cao, H. Shindo, J. Xue, Q. Delfosse, K. Kersting, STORM: Segment, track, and object re-localization from a single image, 2025, http://dx.doi.org/10.48550/ARXIV.2511.09771.

[204] Y. Zheng, Y. Zhang, J. Yu, W. Guo, Y. Xie, Terrain-aware hierarchical control framework for dynamic locomotion of humanoid robots, Electronics 14 (7) (2025) 1264, http://dx.doi.org/10.3390/electronics14071264.

[205] A.R. Ali, H. Kamal, Time-to-fault prediction framework for automated manufacturing in humanoid robotics using deep learning, Technologies 13 (2) (2025) 42, http://dx.doi.org/10.3390/technologies13020042.

[206] J. Jayawardana, H. Dilshan, R. Wijethilaka, T. Balasooriya, U. Rajapaksha, S. Harshanath, C. Jayawardena, Train a robot to climb staircase using vision-base system, in: 2022 IEEE 10th Region 10 Humanitarian Technology Conference (R10-HTC), IEEE, Hyderabad, India, 2022, pp. 294–299, http://dx.doi.org/10.1109/R10-HTC54060.2022.9930098.

[207] S. Armleder, Y. Zhang, J.R. Guadarrama-Olvera, G. Cheng, Transporting heavy payloads with a humanoid riding a hoverboard, in: 2025 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), IEEE, Hangzhou, China, 2025, pp. 6587–6592, http://dx.doi.org/10.1109/IROS60139.2025.11247121.

[208] J. Jain, S.V. Iyer, K.S. Sangwan, Dhiraj, Vision-driven robotic grasping order generation using segmentation and relative positioning in a cluttered environment, Procedia Comput. Sci. 258 (2025) 2563–2572, http://dx.doi.org/10.1016/j.procs.2025.04.518.

[209] A. Purushottam, J. Yan, C. Xu, Y. Sim, J. Ramos, Wheeled humanoid bilateral teleoperation with position-force control modes for dynamic loco-manipulation, 2024, http://dx.doi.org/10.48550/ARXIV.2407.12189.