This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

Embodied Intelligence Toward Future Smart Manufacturing in the Era of AI Foundation Model

Lei Ren, *Senior Member, IEEE*, Jiabao Dong, *Graduate Student Member, IEEE*, Shuai Liu, *Student Member, IEEE*, Lin Zhang, *Senior Member, IEEE*, and Lihui Wang

***Abstract*—Embodied intelligence has always been regarded as the ultimate form of artificial intelligence (AI) and an ideal concept for smart manufacturing. With the development of AI foundation models, remarkable generalization capabilities have been achieved in various fields, such as natural language processing and computer vision. In the era of AI foundation models, embodied intelligence will be capable of continuous evolution for unlimited tasks with multimodal physical interaction in the open world. Therefore, it is envisioned that embodied intelligence should be integrated into smart manufacturing to upgrade the industry for more intelligent, flexible, and human-centric manufacturing in the future. Therefore, in this article, the definition and components of embodied intelligence are proposed with its novel characteristics. Besides, the capabilities of embodied intelligence in the era of AI foundation model are discussed. Moreover, typical innovative applications of embodied intelligence throughout the whole product life cycle for smart manufacturing are presented with insights. However, embodied intelligence still faces some challenges for implementation. Thus, in the prospect of future, challenges and outlooks of embodied intelligence are discussed for further research and applications.**

***Index Terms*—Artificial Intelligence (AI) foundation model, embodied intelligence, multimodal model, robotics, smart manufacturing.**

Received 6 June 2024; revised 22 July 2024; accepted 15 August 2024. Recommended by Technical Editor Y. Tang and Senior Editor H. Wang. This work was supported in part by the National Key Research and Development Program of China under Grant 2023YFB3308000, in part by the National Natural Science Foundation of China (NSFC) Project under Grant 62225302, Grant 92167108, and Grant 62173023, and in part by the Academic Excellence Foundation of BUAA for PhD Students. *(Corresponding authors: Lei Ren; Jiabao Dong.)*

Lei Ren is with the School of Automation Science and Electrical Engineering, Beihang University, Beijing 100191, China, and with the State Key Laboratory of Intelligent Manufacturing System Technology, Beijing 100854, China, and also with Zhongguancun Laboratory, Beijing 100094, China (e-mail: renlei@buaa.edu.cn).

Jiabao Dong, Shuai Liu, and Lin Zhang are with the School of Automation Science and Electrical Engineering, Beihang University, Beijing 100191, China, and also with the State Key Laboratory of Intelligent Manufacturing System Technology, Beijing 100854, China (e-mail: aarondong@buaa.edu.cn; liushuai2001@buaa.edu.cn; zhanglin@buaa.edu.cn).

Lihui Wang is with the Department of Production Engineering, KTH Royal Institute of Technology, 10044 Stockholm, Sweden (e-mail: lihui.wang@iip.kth.se).

Color versions of one or more figures in this article are available at https://doi.org/10.1109/TMECH.2024.3456250.

Digital Object Identifier 10.1109/TMECH.2024.3456250

I. Introduction

THE concept of embodied intelligence was first proposed by Alan Turing in the early 1950s [1], referring to machines being able to autonomously interact with the environment, such as humans, perceive, decide, plan, act, and possess executive capabilities [2]. Embodied intelligence is considered as the ultimate form of artificial intelligence (AI).

With the development of information and communication technologies represented by AI, some efforts have been made to deploy AI models onto machines [3], [4]. These contributions aim to apply the concept of embodied intelligence in areas, such as home care services [5], [6], smart vehicles [7], and particularly in robotics of industrial production and smart manufacturing [8]. Recently, plenty of works have been exploring to realize embodied intelligence and enhance smart manufacturing via AI methods [9], [10]. Nevertheless, despite the progress made by AI in smart manufacturing, there are still certain challenges to be addressed toward future smart manufacturing [11]. The future smart manufacturing aims to enable mass personalization and highly flexible production efficiently in response to changing consumers' demands and producing plans via multimodal interaction. In contrast, existing AI methods in smart manufacturing have limitations as follows.

1) *Limited manufacturing tasks:* Existing AI methods fall short in handling unlimited customized manufacturing tasks for mass personalization. Because they are typically designed for specific and predefined tasks, they lack the flexibility to adapt to highly variable and unique requirements [12].
2) *Structured manufacturing environments:* Existing AI methods often fail to generalize in flexible, changing, and unstructured manufacturing environments. Due to their reliance on fixed models and predefined rules, these methods lack the adaptability to handle unforeseen scenarios and unpredictable settings.
3) *Inefficient human–machine interaction:* The manufacturing human–machine interaction still lacks effective collaboration. The control manner for manufacturing machines is still inefficient, mainly limited to teaching and coding, which is labor-intensive and requires a high level of expertise.
4) *Lack of collaborative intelligence:* In the industrial internet, a series of tasks in production line are usually continuous with common knowledge and mechanisms.

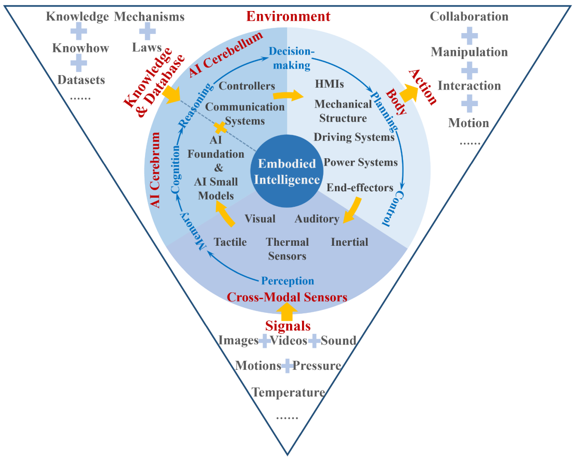


Fig. 1. Embodied intelligence model.

However, multiple machines in production line have independent small AI models operating without sharing intelligence due to limited interoperability and isolated data silos among models.

5) *Single-input processing limitation:* Most of existing AI models can only process and respond to one type of input at a time, either internal, external, or task-specific information, rather than integrating and responding to all simultaneously in an unified framework.

With the advance of the era of AI foundation models, extraordinary generalization capabilities have been achieved in various domains such as natural language processing [13] and computer vision [14] through learning from massive data. Therefore, it is envisioned that the concept of embodied intelligence will be expanded with AI foundation models to tackle abovementioned challenges. In the era of AI foundation model, the embodied intelligence is defined as an intelligent system that interacts with the external open world to acquire information [15], understand the world, make decisions, and take actions to influence the world physically. Such systems consist of AI cerebrum, cerebellum, body, and cross-modal sensory system, which is named an ABC embodied intelligence structure in this article, as illustrated in Fig. 1. The AI cerebrum and cerebellum jointly conduct cognition, reasoning, and decision-making based on perception and memory. The body plans and controls to act based on decision instruction from the cerebrum and cerebellum [16]. The cross-modal sensory system perceives multimodal information from both external environment and internal body. Furthermore, the ABC embodied intelligence structure are proactively learning from data collected during operation, enabling continuous learning and evolution [17] to generate intelligent behaviors with adaptability [18].

With the proposed concept and structure, the embodied intelligence is capable of continuous learning and evolution for unlimited tasks and generalized in dynamic environments of open world [19]. In addition, the embodied intelligence can realize multimodal interaction with human and influence upon real world physically. Besides, the entities of embodied intelligence can be multiple machines, which together form series connection or bottom-up collaborative intelligence as production lines, factories, and even supply and industrial chains. Moreover, the embodied intelligence processes proprioception, exteroception, and task-specific input simultaneously in a unified framework [20], [21].

Therefore, with the abovementioned characteristics and potential capabilities, it is necessary to integrate embodied intelligence into future smart manufacturing to address abovementioned issues. To the best of authors' knowledge, this is the first work discussing embodied intelligence toward future smart manufacturing. In this article, the definition and model of embodied intelligence is proposed with analysis of its components and inner operation logic. Then, the novel characteristics and capabilities of embodied intelligence in the era of AI foundation model are discussed, respectively. In addition, the technical architecture of embodied intelligence system for smart manufacturing is presented. Moreover, typical innovative applications of embodied intelligence throughout the product life cycle in smart manufacturing are presented. Finally, in spite of the ideal concept of embodied intelligence, there are still challenges and open issues to discuss, and therefore, outlooks are presented for further relevant research and applications.

II. Typical Characteristics of Embodied Intelligence in the Era of AI Foundation Model

In the era of AI foundation model, embodied intelligence is featured with novel characteristics, which will address challenges summarized in Section I toward future smart manufacturing.

A. Continuous Learning and Evolution for Unlimited Tasks

Existing AI methods are designed for limited, specific, and predefined tasks, lacking the flexibility adapting to highly variable and unique requirements. They can only passively accept human-collected or curated data rather than autonomously learning [22].

To realize mass personalization, the ability catering to diverse individual preferences and requirements is needed. Embodied intelligence needs adaption to various production tasks over time, which is realized in a pipeline of automatic data collection and continuous learning and evolution. Embodied intelligence should learn more complex tasks and unseen objects from previous simpler ones, such as learning more complex assembly task from easier grasping task, or learning to handle unseen objects. This means extensive customization options will be provided while maintaining efficiency and scalability in production processes.

B. Generalized in Dynamic Environments of Open World

Current manufacturing processes are constrained by structured manufacturing environments. While changing production demands, adjustment of the production line is necessary. Existing AI models often fail to generalize in working in flexible, changing, and unstructured manufacturing environments.

Toward future smart manufacturing, the generalization in dynamic environments is envisioned [23]. Embodied intelligence presents a solution to the challenges posed by the changing open world. Embodied intelligence generalizes in flexible, changing, and unstructured manufacturing environments by leveraging the generalization capabilities of AI foundation models. It adapts to varying environments and conditions, learning from simple to complex scenarios. This adaptability ensures an efficient performance and resilience for dynamic industrial settings, even facing unexpected disruptions.

C. Physical Multimodal Interaction With Real World

The state-of-the-art AI systems, such as ChatGPT and Sora, referred to as disembodied AI, still mainly exist in the digital domain, lacking physical bodies. Therefore, they cannot exert influence on the physical world without the embodiment [24]. While these cutting-edge large multimodal models exhibit remarkable capabilities, their ability to generalize across interactive contexts remains limited.

The ultimate form of artificial general intelligence (AGI) is envisioned to be embodied in machines equipped with physical bodies, such as robots. It is anticipated that in the future, AGI will be applied to change, influence, and interact with the physical world [25]. Embodied intelligence enables collaborative interaction with humans to accomplish various tasks by leveraging the interactive capabilities given by large multimodal models with embodiment [26]. Therefore, toward future smart manufacturing, embodied intelligence will assist humans in completing tasks via multimodal interaction in the physical world.

D. Collaborative Swarm Intelligence

In the current industrial Internet of Things (IIoT), devices only have simple connections within distributed swarm intelligence systems. For instance, during a series of tasks in one production line, tasks such as robotic grasping and assembly involve objects with similar characteristics, but independent AI models for these tasks cannot share parameters to enhance effect together [27].

The future entails the collaboration of multiple machines, allowing for the sharing of contextual semantic information, knowledge, and understanding across different production tasks. In the system of the IIoT, various devices will be equipped with embodied intelligent agents, sharing intelligence, and working together to accomplish tasks. This collaborative swarm intelligence helps form a bottom-up embodied intelligence system, including machines, production lines, factories, and supply and industrial chains from bottom up, which will be further discussed in Section III.

E. Proprioception, Exteroception, and Task-Specific Action

Current systems, whether driven by nonintelligent or intelligent models, often rely on singular modes of response. Devices lack the ability to both perceive environment and recognize their own bodies, leading to limitations in their adaptability and responsiveness.

The future involves the integration of diverse forms of signals, encompassing proprioception, exteroception, and task-specific action, to enable both rapid instinctual responses and intelligent reasoning decisions. Embodied intelligence entails simultaneous processing of proprioceptive, exteroceptive, and task-oriented signals, akin to the collaborative mechanism between the human cerebrum and cerebellum. The cerebrum receives task-oriented signals, responsible for complex logical reasoning tasks. Conversely, the cerebellum processes proprioceptive signals to control different parts of the body and responds instinctively to changes in the external environment [28]. By combining task instructions and considering both rapid responses and decisions, tasks will be efficiently completed in future smart manufacturing.

III. Capabilities of Embodied Intelligence in Industrial Units for Smart Manufacturing

In this section, capabilities will be analyzed in dimension of intelligent manufacturing units as illustrated in Fig. 2. These units include machine, production line, factory, and supply and industrial chain from individual in basic level to the whole in higher level. The entities of embodied intelligence can be robots with various types, machine tools, production equipment, and hardware products. The entities can also be expanded to higher levels, such as combinations of multiple embodied intelligent entities forming production lines, factories, supply, and industrial chains [29]. This forms a bottom-up embodied intelligence collaboration in the smart manufacturing.

A. Machine

Perception of machine: Machines can perceive multimodal information in the production environment, such as vision, language, pressure, and temperature, which are collected to be analyzed by multimodal foundation models.

Memory of machine: Machines are capable of storing and remembering multimodal information related to production. By compressing information from recognized patterns of data, machines can complete different tasks throughout the production process.

Cognition of machine: The multimodal foundation model of the machine integrates and correlates the perceived and memorized multimodal information, thereby gaining a comprehensive understanding of the production environment. This allows for a better simulation of the human cognitive process and provides support in complex and dynamic industrial environments.

Reasoning of machine: The multimodal embodied intelligence reasons from the results of cognition to deduce the overall production process and the logical relationships among different production steps, ensuring that the production procedures are carried out in sequence.

Decision-making of machine: Based on the results of reasoning, the embodied intelligence can automatically determine the current production phase of the machine and accordingly make next decision of the production process.

Planning of machine: Based on the comprehensive results of reasoning and decision-making, the embodied intelligence will

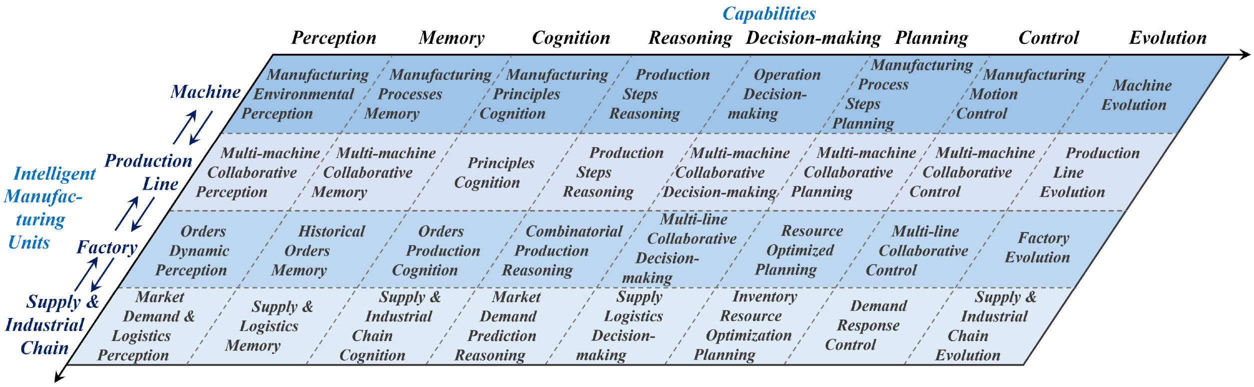


Fig. 2. Capabilities of embodied intelligence in industrial units for smart manufacturing.

further plan and select appropriate methods to advance the production process in accordance with the outcomes of reasoning and decision-making, ensuring further smooth control.

Control of machine: After planning, the embodied intelligence will use actuators, controllers, and communicators to precisely control the machines [30], ensuring production activities carried out according to the predetermined procedures and in the predefined manner.

Evolution of machine: During the execution of tasks, the embodied intelligence not only continuously perceives, reasons, and makes decisions to ensure the smooth progress of production tasks, but also gathers data throughout the production process, expanding its database, and thereby gradually evolving its learning capabilities.

B. Production Line

Perception of production line: On the production line, perception is achieved through sensors distributed across various machines. The production line collects multimodal information such as visuals data from quality control device, and tactile data from dexterous hands. These multimodal data are perceived by various machines, and then transmitted with comprehensive information to the embodied intelligence of production line for further processing and analysis.

Memory of production line: The production line summarizes and stores the multimodal information related to production that machines remember and store during the production process. Utilizing this comprehensive compressed information, the production line will complete the manufacturing tasks of the products with prior memory.

Cognition of production line: The multimodal embodied intelligence integrates and correlates the multimodal information perceived by each machine at various stages of production, achieving an in-depth understanding of the entire production process and environment. This provides precise cognitive support and operational guidance for industrial production.

Reasoning of production line: The embodied intelligence in the production line can reason the overall production process of the product. The production processes of different machines will also be coordinated based on the cognitive information in various production steps.

Decision-making of production line: Based on the information from reasoning, the embodied intelligence can autonomously determine the current status of the production line and decide the next processing direction for the product, ensuring that the production line operates according to the reasoning production process.

Planning of production line: After determining the product processing route, the embodied intelligence precisely selects the appropriate machines to execute the subsequent processing steps based on the real-time production status of each machine within the production line.

Control of production line: The control of production line not only involves the precise regulation of the machine production process using various controllers and actuators, but also requires intelligent control over the positioning and production paths of machines within the line.

Evolution of production line: Throughout the production process, the embodied intelligence continuously evolves itself through activities of perception, cognition, reasoning, etc. This achieves autonomous evolution, enhancing performance, and capabilities to meet dynamic and flexible production task requirements, ensuring the stability and efficiency of the production process.

C. Factory

Perception of factory: In a factory, perception is carried out through various sensors among product production lines. Information about the entire factory's production environment is perceived and aggregated. This comprehensive information is ultimately transmitted to the multimodal embodied intelligence for subsequent memory and cognition.

Memory of factory: During the product production, information is collected by the production lines, such as machine operational data, product quality indicators, and environmental conditions. This comprehensive information is summarized and stored by the factory's intelligent agent system, which helps coordinating the production line to complete different production tasks.

Cognition of factory: The embodied intelligence in the factory is capable of integrating information from each independent production line, forming a comprehensive understanding of the entire factory's production process and environment. This enables the production of factory with enhanced efficiency and adaptability.

Reasoning of factory: The embodied intelligence can integrate information from various production lines, reason about the overall production process of the factory, and the coordination of product production among different lines. Based on reasoning capability, the production plan for the entire factory can be formulated, ensuring the stability of the factory's production process.

Decision-making of factory: The embodied intelligence will autonomously determine the current stage of production and decide the next production direction based on the reasoning results. It will recognize the production status of the factory and respond quickly and accurately to any production demands, then coordinate the production lines to proceed manufacturing process.

Planning of factory: After making decisions, the embodied intelligence will further plan by scheduling and collaborating the appropriate production lines based on the current factory production situation to complete the production plan.

Control of factory: The factory will control the operation of different production lines according to the established decisions and plans, ensuring the stable operation of the production process.

Evolution of factory: The embodied intelligence is capable of monitoring changes in the production environment and production requirements in real time during the production process and adjusting the production strategy accordingly. As the data becomes more abundant for evolution, the learning capability of the embodied intelligence will be further enhanced, enabling autonomous learning and development.

D. Supply and Industrial Chain

Perception of supply and industrial chain: Perception is responsible for collecting and aggregating market demand information, production information, and other critical supply and industrial chain data. After integration, this information is transmitted to the multimodal embodied intelligence, providing comprehensive perception support.

Memory of supply and industrial chain: The supply and industrial chain system stores and compress the collected and aggregated information for further data processing and analysis, ensuring the stable operation and continuous improvement of the supply and industrial chain system.

Cognition of supply and industrial chain: The embodied intelligence within the supply and industrial chain integrates multisource data such as production processes and market demand to conduct in-depth analysis. This forms a comprehensive understanding of factory operations and market conditions, enhancing its adaptability to market changes.

Reasoning of supply and industrial chain: The supply and industrial chain system utilizes its reasoning capabilities to predict future supply and production needs based on the real-time production status of the factory and current market demand.

Decision-making of supply and industrial chain: The embodied intelligence will consider the market conditions, production status, and reasoning results comprehensively to determine the

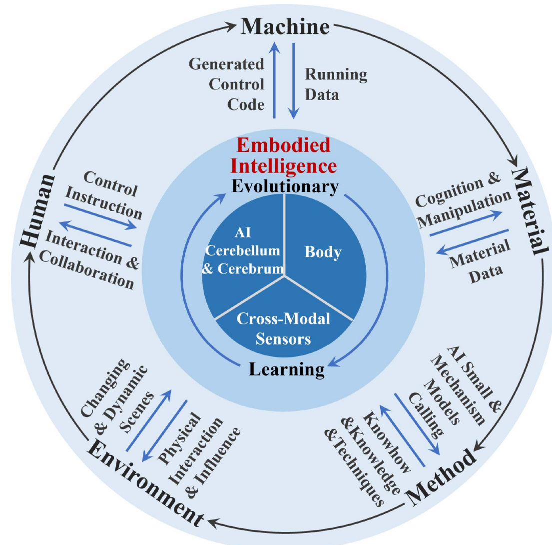


Fig. 3. Interaction between embodied intelligence and industrial elements.

supply of production materials and the manufacturing of products, ensuring a balance of supply and demand in the production process.

Planning of supply and industrial chain: Based on the results of reasoning and decision-making, the embodied intelligence will adjust the production and material supply of the entire factory, coordinating the production among the factories to ensure production balance.

Control of supply and industrial chain: The embodied intelligence will precisely control production and supply–demand based on the content of decisions and plans, ensuring the smooth execution of planned tasks. This achieves efficient and stable operation within the supply and industrial chain.

Evolution of supply and industrial chain: The embodied intelligence of the supply and industrial chain monitors the production status, market environment, and changes in the supply–demand structure in real time. It adjusts production strategies based on these changes to achieve adaptive production management. Simultaneously, the data will be continuously enriched during the adjustment process, further promoting the evolution of the supply and industrial chain.

IV. Architecture of Industrial Embodied Intelligence System

In the industrial embodied intelligence system, the interaction relationship between embodied intelligence with industrial elements is illustrated by Fig. 3 and described as follows.

1) *Human:* Embodied intelligence receives multimodal interaction commands from humans and understands intentions and needs, then collaborate with human to complete production tasks according to the commands.
2) *Machine:* Embodied intelligence interacts with machines through commands to control their operation, in which sensors collect machine running data, which is used by the embodied intelligent cerebellum to adjust its working mode.

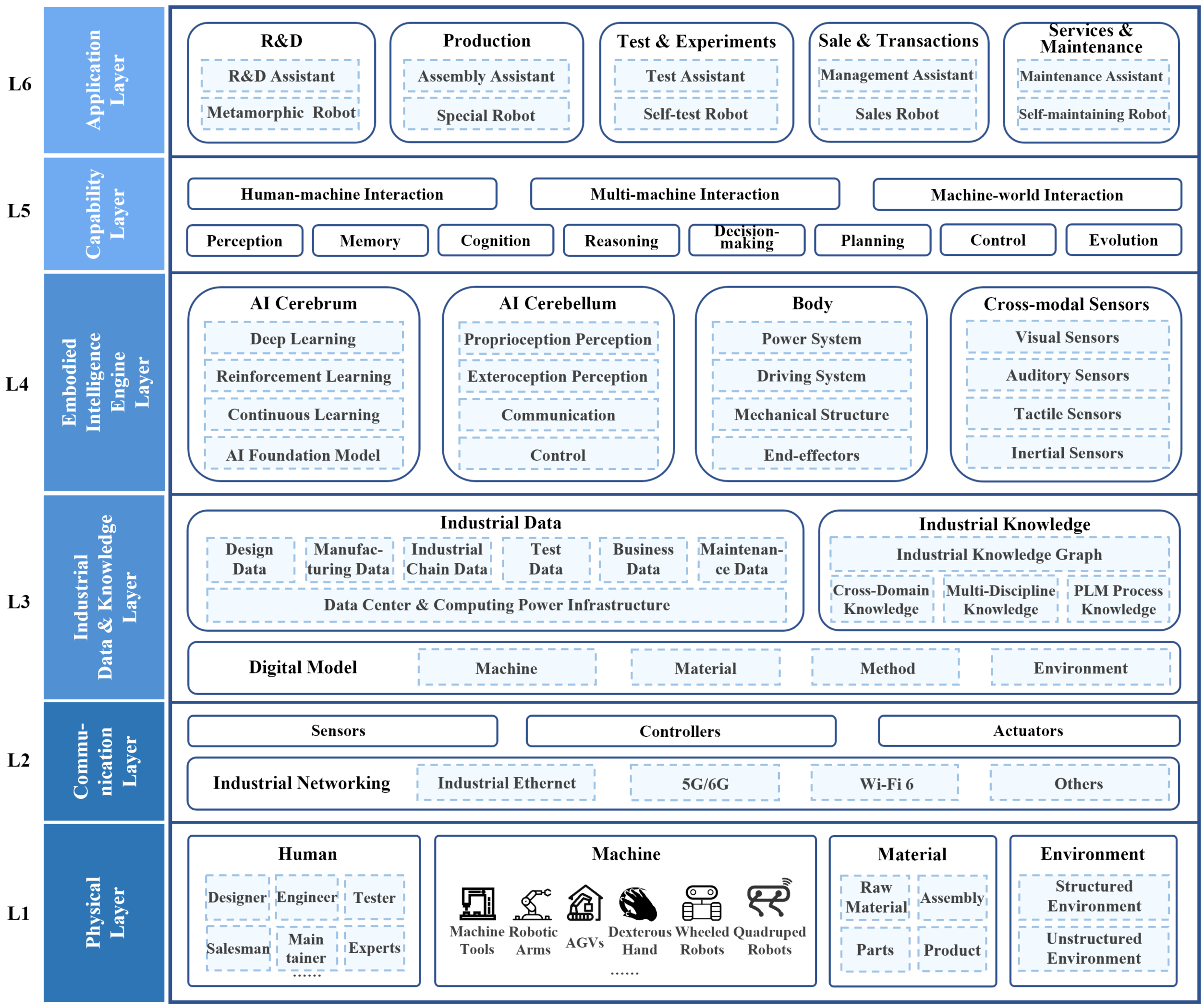


Fig. 4. Technical architecture of embodied intelligence system.

3) *Material:* Embodied intelligence perceives and manipulates materials for production and manufacturing, and materials provide data for the embodied intelligent cerebrum for cognition and reasoning in production tasks.
4) *Method:* Embodied intelligence invokes external models or methods for production tasks, which provide the embodied intelligence with the knowledge, knowhow, and techniques needed in manufacturing.
5) *Environment:* Embodied intelligence senses environmental information and adjust its operation strategy accordingly. The challenges brought by unstructured environmental changes in open world also promote the continuous evolution of embodied intelligence.

Moreover, the technical architecture of embodied intelligence system, described by Fig. 4, will be discussed as follows.

A. Physical Layer

In the technological architecture of industrial embodied intelligence, the physical layer serves as the foundation, encompassing four key components: human, machine, material, and environment. The human element represents the engineers, designers and other people involved in manufacturing processes. Machines refer to the various robotic and automated systems utilized in manufacturing and production, ranging from robotic arms, various robots to machine tools. Materials include the raw material, parts, and finished products manipulated and processed within the manufacturing process. The environment includes the physical surroundings where industrial activities take place, mainly divided into structured and unstructured environments. Together, these components form the physical layer, which is fundamental to support various activities in smart manufacturing.

B. Communication Layer

The communication layer represents an essential interface facilitating seamless interaction and data exchange between various components within the smart manufacturing system. This layer comprises three primary elements: sensors, actuators, and controllers. Sensors play a crucial role in capturing real-time data from the physical environment, including vision,

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

auditory, pressure, temperature, etc., providing valuable inputs for monitoring and control systems. Controllers serve as the center, organizing, and regulating the actions of machines and processes based on inputs from sensors and directives from higher level control systems. Actuators convert electrical energy into mechanical movement by manipulating mechanisms, in order to interact with the external environment, and provide feedback to the system, allowing for closed-loop control and adjustment based on real-time performance. Together, these elements form a robust communication infrastructure that enables the efficient transmission of information, commands, and feedback, facilitating agile and adaptive operations for various tasks of embodied intelligence.

C. Industrial Data and Knowledge Layer

The industrial data and knowledge layer serves as a crucial repository and processing hub for industrial data and knowledge, as well as the digital models established for embodied intelligent entities. This layer comprises two primary components: industrial data and knowledge, and digital models for embodied intelligent entities. Industrial data encompasses data collected from sensors, machines, processes, and human interactions throughout the product life cycle. These data are collected, stored, and processed to derive insights, patterns, and trends that inform decision-making and optimization efforts in smart manufacturing. Industrial knowledge includes various knowledge with specific industrial expertise relevant to industrial operations. Digital models for embodied intelligent entities represent virtual representations of physical entities within the smart manufacturing, facilitating simulation, analysis, and optimization of their behavior and interactions. Together, these components form a robust foundation for leveraging data-driven insights and digital representations to enhance productivity, efficiency, and innovation in industrial processes.

D. Embodied Intelligence Engine Layer

The embodied intelligence engine layer constitutes a critical component responsible for orchestrating the cognitive and sensory functions of embodied intelligence. This layer comprises four primary elements: AI cerebrum, AI cerebellum, body, and cross-modal sensors. The AI cerebrum serves as the central processing unit, responsible for higher level cognitive functions, such as reasoning, planning, and decision-making within the industrial context. The AI cerebellum, akin to its biological counterpart, coordinates sensorimotor functions, integrating feedback from cross-modal sensors to regulate and refine the movements and interactions of the embodied intelligence with environment. The body refers to the physical structure and actuators of the system, enabling it to manipulate objects, navigate spaces, and execute tasks autonomously. Cross-modal sensors capture a diverse range of sensory inputs, including exteroceptive visual, auditory, tactile data, and proprioceptive information, providing the system with comprehensive situational awareness and facilitating adaptive behavior in dynamic industrial environments [31]. Together, these components collaborate as engines in embodied intelligence to operate effectively and autonomously.

E. Capability Layer

The capability layer constitutes the cognitive and functional capabilities of embodied intelligence. These capabilities include perception, memory, cognition, reasoning, decision-making, planning, control, and evolution, which are described in detail in Section III. Aside from these capabilities, interaction is also a significant capability in higher level, which calls these capabilities above. The interaction can be mainly divided into human–machine interaction, multimachine interaction, and machine-world interaction. These interaction capabilities facilitate seamless collaboration in diverse and dynamic industrial environments of smart manufacturing.

F. Application Layer

In the industrial embodied intelligence architecture, the application layer represents the practical implementation and deployment of embodied intelligence across various stages of the industrial life cycle. These applications will be further discussed in next section in dimension of product life cycle and human involvement together.

V. Typical Applications of Embodied Intelligence for Smart Manufacturing

With remarkable capabilities of the embodied intelligence, it is prospected that its applications will cover the whole product life cycle. In this article, the product life cycle is divided into research and development (R&D), production, test and experiments, sales and management, and services and maintenance. In each phase of product life cycle, according to human involvement, we propose separate typical applications for human-centric and embodied autonomous applications, respectively. These applications are illustrated in Fig. 5.

A. Research and Development

R&D assistant: R&D assistants collect and analyze performance data of existing products as well as user feedback to provide suggestions for designers to design products. During the R&D design phase, wearable devices can serve as physical embodiment of intelligent design assistants. Through embodied intelligent wearable devices, R&D assistants help designers via render prototypes for visualization, simulate assembly processes, and adjust design parameters from human instructions. During the assembly process of prototypes, wearable devices can monitor performance parameters in real time, which helps designers understand the operating status of prototypes and promptly identify potential issues intuitively and efficiently, thus enhancing design quality, shortening R&D cycles, and achieving higher quality and innovative designs.

Metamorphic robot: Embodied intelligence can also achieve autonomous optimized design. Metamorphic robots are robots capable of bodily, modal, and functional changes and evolution, which exemplify autonomous evolutionary capabilities of embodied intelligence [32]. They can autonomously change structure, morphology, and modalities and evolve in real-time according to the environment. The evolving targets can be structure and

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

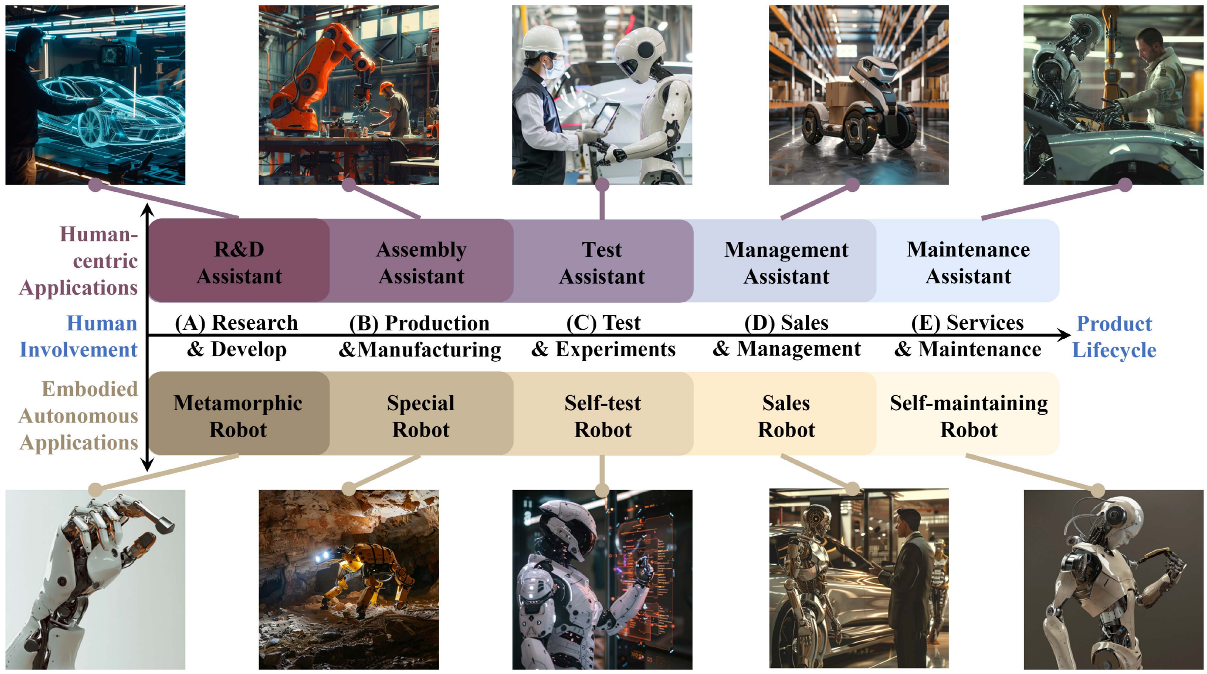


Fig. 5. Typical applications of embodied intelligence for smart manufacturing in product life cycle.

material of embodied entities. For instance, in the design process of robotic dexterous hands, metamorphic robot collect data on hand movements and object grasping to analyze the optimal grasping strategies, then adjust the shape or structure of the hands to improve grasping efficiency and adaptability. For quadruped robots working in outdoor environments, the efficiency of robot movement on different terrains will be analyzed by monitoring joint loads and motion data, and adjusting joint stiffness and damping.

B. Production and Manufacturing

Assembly assistant: Embodied intelligence-based assembly assistant collaborates with workers to complete assembly tasks for mass personalization. These robots assist workers in the assembly process by providing functions such as assembly position prediction and error warnings. Assembly assistant collaborate with workers in completing tasks that are too labor-intensive, thereby reducing the physical burden on workers. This allows workers to focus on more creative tasks, such as optimizing assembly processes and inspecting assembly results, thereby improving overall work efficiency and product quality.

Special robot: In certain special environments, special robots replace workers to autonomously perform production tasks, ensuring worker safety and improving operational efficiency. In tasks such as welding, special robots are designed to complete welding tasks autonomously, thereby avoiding workers' direct exposure to high temperatures and radiation. In the mining industry, special robots perform tasks such as underground excavation, transportation, and exploration, avoiding miners working in narrow, dark, and hazardous environments. In underwater construction, salvage, or marine resource exploration, special robots perform precise underwater welding, cutting, or equipment installation tasks. In aerospace, assembly robots autonomously identify, disassemble or assemble components of satellites or space stations in space.

C. Experiment and Testing

Test assistant: Test assistants help testers in completing experimental and testing tasks as physical embodiment. Test assistants perform mechanical tasks such as experimental environment construction and equipment handling before experiments to reduce repetitive work of humans. During the experiments, test robots perceive real-time key indicators such as operating efficiency, load capacity, battery life, and obstacle avoidance ability of products. Ultimately, the test assistants automatically generate test reports containing various key indicators, charts, analytical results, and improvement suggestions for testers to quickly understand the test results.

Self-test robot: During the testing process, self-test robots not only collect and monitor their own real-time parameters, but also conduct in-depth analysis of the test environment and adjust testing conditions to ensure the smooth progress of the test. In addition, the self-test robots independently identify potential issues, analyze the root causes of problems, thereby optimizing the entire testing process. Based on accumulated knowledge and experience, the self-test robots provide analytical results and targeted optimization suggestions, which helps identifying potential shortcomings and improving themselves.

D. Operations and Management

Management assistant: Management assistant can monitor production chains in real time, collect, and summarize

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

production data then provide intelligent suggestions for managers. Based on operation data and management knowledge, management assistants provide future demands prediction, market dynamics, etc., assisting managers in understanding and achieving comprehensive monitoring and optimization of the supply and industrial chains. This helps improving production efficiency, reduce costs, cater to market supply and demand, and enhance the competitiveness of enterprises.

Sales robot: Sales robots autonomously use their multimodal interaction capabilities to provide intelligent sales services for customers. Sales robots interact with customers through voice, touchscreen, or gesture recognition, providing product information, answering questions, and even consulting on returns and exchanges, guiding customers through visits. Based on customers' purchase history or real-time interaction data, sale robots can accurately recommend related products, thereby increasing sales opportunities.

E. Service and Maintenance

Maintenance assistant: Maintenance assistants, especially in complex or hazardous environments, can assist human monitor and inspect to ensure the safety of maintainers. With intelligent reasoning capability, maintenance assistant helps locate problems more quickly and provide corresponding solutions. Afterwards, professional diagnostic suggestions will be provided to workers, and some difficult maintenance tasks, such as heavy object handling and fixture adjustment, will be precisely performed by assistants. In this way, robots can assist maintenance personnel in completing complex maintenance tasks, reducing their physical burden, and enabling them to focus on fine maintenance tasks and other high-skilled processes.

Self-maintaining robot: Self-maintaining robots are embodied intelligent robots with strong autonomous maintenance capabilities. Self-maintaining robot can perform predictive maintenance based on real-time data parameters itself, adjust its own operation modes in a timely manner to minimize damage via AI cerebellum, thereby extending the service life of the equipment and improving production efficiency in optimal condition. In addition, robots can automatically perform maintenance tasks, such as lubrication, cleaning, and replacing worn components, reducing the need for manual intervention.

VI. Challenges and Outlooks

A. Multimodal Data Scaling-Up for AI Foundation Model Learning

Embodied intelligence faces significant challenges in scaling up multimodal data, which is crucial for enhancing cognitive and operational capabilities. For embodied intelligence, multimodal data provide a comprehensive understanding of the environment, especially in open world. It is necessary to scale up multimodal data to support training AI foundation models for embodied intelligence. However, there is still a lack of such real-world multimodal 3-D dataset for embodied intelligence at present. Compared to large-scale 2-D vision-language dataset like contrastive language-image pretraining [33] or segment anything-1B [34], data in 3-D world are more difficult to obtain and annotate. Data for embodied intelligence are heterogeneous from multiple sources to collect and integrate, and multimodal in diverse labels to annotate. Addressing these challenges involves developing more automated data processing and annotation methods and innovative multimodal data fusion alignment methods to enable training better embodied intelligence.

B. Self-Evolution Framework Through AI Foundation Model

Embodied intelligence encounters significant challenges in implementing a self-evolution framework, crucial for continuous improvement, and adaptation in dynamic environments. This framework involves the autonomous collection of data during operation, enabling continuous learning without human intervention. The process begins with the collecting diverse sensory data as it interacts with its surroundings during operation. Then, this data needs to be processed to prepare for training without requiring labor-intensive and time-consuming manual annotation. Techniques such as unsupervised, self-supervised, and reinforcement learning are essential in this context, allowing derivation of meaningful patterns and insights from raw data autonomously. In addition, the framework must ensure efficient data management, including storage, retrieval, and preprocessing, to maintain a high learning rate and adaptability. By overcoming these challenges, embodied intelligence enhances its cognitive flexibility, capable of learning unlimited tasks in dynamic real-world scenarios.

C. Real-Time Human-Interaction Via AI Foundation Model

Embodied intelligence faces substantial challenges in achieving effective real-time human interaction, which is crucial for human-centric applications in smart manufacturing. Real-time interaction necessitates multimodal capabilities, enabling processing and reasoning across diverse multimodal data streams, including 3-D vision, audio, text, tactile signals, etc. The main solution is multimodal alignment to deal with in AI foundation models, which helps understand spoken instructions, interpret visual cues from the surrounding workspace, and respond appropriately to textual information. Another issue to tackle with is synchronization and low-latency processing of these multimodal data to provide timely and contextually relevant responses with high frequency, which requires the integration of sophisticated multimodal alignment to ensure seamless interaction. Moreover, the robots must be capable of adapting to dynamic and unpredictable human behaviors while maintaining safety and efficiency, enabling robots to interact naturally and effectively with human.

D. Industrial Knowledge-Based Embodied Intelligence

The industrial knowledge integration is a significant challenge for embodied intelligence particularly in the era of AI foundation model, which may output untrustworthy results and cause hallucinations. Embedding industrial knowledge involves synthesizing domain-specific information, including manufacturing processes, safety protocols, quality standards, etc., into the cognitive framework of embodied intelligence. This integration ensures that decision-making processes align with industrial best practices and regulatory requirements, mitigating the risk of erroneous or unsafe behavior in manufacturing tasks. Furthermore, to prevent hallucinations and uncertainties of AI foundation models, robust validation, and verification mechanisms are essential. Techniques like uncertainty quantification, model calibration, and anomaly detection help enhance the reliability and trustworthiness of predictions and embodied actions.

E. Reasoning Replacing Compression in AI Foundation Model

The future prospect of embodied intelligence with AI foundation models involves enhancing its reasoning capabilities rather than merely compressing massive information. This enables embodied intelligence to delve deeper into understanding patterns behind massive data and infer more complex, higher level relationships and structures. Reasoning involves making deductions, inferences, and speculations about data rather than simply conducting straightforward statistical analysis or pattern matching. Through reasoning, embodied intelligence handles unseen, incomplete or ambiguous information better, and makes more accurate decisions and planning. The development of this capability is significant to understand and deal with the complexity and uncertainty in many fields especially in changing environments of flexible manufacturing. Moreover, as embodied intelligence continues to learn and evolve, the model will become more creative and innovative. With reasoning capability in the open world, embodied intelligence will generate new ideas, designs, and solutions, driving progress in R&D, industrial process optimization, technique innovation, and other applications.

F. Advanced Action System in Brain of Embodied Intelligence

Embodied intelligence faces the challenge of developing an advanced action system that mirrors the complex interplay between the cerebrum and cerebellum of humans. In this intricate neural system, the cerebrum handles cognition and reasoning, responsible for higher order intelligent tasks. Conversely, the cerebellum integrates sensory inputs from the external environment and proprioceptive feedback from its own body, coordinating the execution of instinctual and motor functions. This system will help both cognitive reasoning for complex tasks and real-time motor control for embodied action, which correspond to understanding abstract instructions from human and basic obstacle avoidance, respectively. With synergy of these two action mechanisms, embodied intelligence will be enabled with human-like adaptability, versatility, and autonomy in dynamic environments.

VII. Conclusion

In the era of AI foundation models, AI has achieved excellent generalization in multiple tasks and fields. With prevailing trend of the AI foundation model, the integration of AI foundation model and embodied intelligence is promising. To the best of authors' knowledge, as the first work discussing embodied intelligence toward future smart manufacturing, we envisioned that the embodied intelligence will bring next-generation smart manufacturing enhancement and more opportunities. Therefore, the definition of embodied intelligence and its corresponding novel characteristics were proposed. Compared with previous AI applications in smart manufacturing, the novel capabilities of embodied intelligence are proposed in hierarchical industrial units. Moreover, the embodied intelligence's typical innovative applications were discussed with insights for future usage throughout product life cycle. Finally, challenges and outlooks were discussed as future research directions.

References

[1] A. Zador et al., "Catalyzing next-generation artificial intelligence through neuroai," *Nature Commun.*, vol. 14, no. 1, 2023, Art. no. 1597.

[2] T. Wang, P. Zheng, S. Li, and L. Wang, "Multimodal human–robot interaction for human-centric smart manufacturing: A survey," *Adv. Intell. Syst.*, vol. 6, no. 3, 2024, Art. no. 2300359.

[3] J. Yang, X. Wang, Y. Tian, and F.-Y. Wang, "Parallel intelligence in CPSSs: Being, becoming, and believing," *IEEE Intell. Syst.*, vol. 38, no. 6, pp. 75–80, Nov./Dec. 2023.

[4] E. Ottonello, M. Baggetta, G. Berselli, and A. Parmiggiani, "Design and validation of a push-latch gripper made in additive manufacturing," *IEEE/ASME Trans. Mechatron.*, vol. 28, no. 4, pp. 2083–2091, Aug. 2023.

[5] K. Fan, M. Jouaiti, A. Noormohammadi-As, C. L. Nehaniv, and K. Dautenhahn, "A social referencing disambiguation framework for domestic service robots," in *Proc. IEEE Int. Conf. Robot. Autom.*, 2023, pp. 11979–11985.

[6] Y. Liu et al., "Multimodal embodied interactive agent for cafe scene," 2024, *arXiv:2402.00290*.

[7] S. Sujitha, S. Pyari, W. Jhansipriya, Y. R. Reddy, R. V. Kumar, and P. R. Nandan, "Artificial intelligence based self-driving car using robotic model," in *Proc. 3rd Int. Conf. Artif. Intell. Smart Energy*, 2023, pp. 1634–1638.

[8] L. Ren, J. Dong, D. Huang, and J. Lü, "Digital twin robotic system with continuous learning for grasp detection in variable scenes," *IEEE Trans. Ind. Electron.*, vol. 71, no. 7, pp. 7653–7663, Jul. 2024.

[9] L. Ren, J. Dong, X. Wang, Z. Meng, L. Zhao, and M. J. Deen, "A data-driven auto-CNN-LSTM prediction model for lithium-ion battery remaining useful life," *IEEE Trans. Ind. Informat.*, vol. 17, no. 5, pp. 3478–3487, May 2021.

[10] G. Bravo-Palacios, H. Li, and P. M. Wensing, "Engineering compliance in legged robots via robust co-design," *IEEE/ASME Trans. Mechatron.*, to be published, doi: 10.1109/TMECH.2024.3382305.

[11] L. Ren et al., "Industrial metaverse for smart manufacturing: Model, architecture, and applications," *IEEE Trans. Cybern.*, vol. 54, no. 5, pp. 2683–2695, May 2024.

[12] Y. Long, W. Wei, T. Huang, Y. Wang, and Q. Dou, "Human-in-the-loop embodied intelligence with interactive simulation environment for surgical robot learning," *IEEE Robot. Autom. Lett.*, vol. 8, no. 8, pp. 4441–4448, Aug. 2023.

[13] Z. Xi et al., "The rise and potential of large language model based agents: A survey," 2023, *arXiv:2309.07864*.

[14] D. Zhu, J. Chen, X. Shen, X. Li, and M. Elhoseiny, "MiniGPT-4: Enhancing vision-language understanding with advanced large language models," 2023, *arXiv:2304.10592*.

[15] X. Wang et al., "Parallel driving with big models and foundation intelligence in cyber-physical-social spaces," *Research*, vol. 7, 2024, Art. no. 0349.

[16] Y. Gan, B. Zhang, J. Shao, Z. Han, A. Li, and X. Dai, "Embodied intelligence: Bionic robot controller integrating environment perception, autonomous planning, and motion control," *IEEE Robot. Autom. Lett.*, vol. 9, no. 5, pp. 4559–4566, May 2024.

[17] A. Gupta, S. Savarese, S. Ganguli, and L. Fei-Fei, "Embodied intelligence via learning and evolution," *Nature Commun.*, vol. 12, no. 1, 2021, Art. no. 5721.

[18] F. Iida and F. Giardina, "On the timescales of embodied intelligence for autonomous adaptive systems," *Annu. Rev. Control, Robot., Auton. Syst.*, vol. 6, pp. 95–122, 2023.

[19] Z. Zhu, X. Ma, Y. Chen, Z. Deng, S. Huang, and Q. Li, "3D-VisTA: Pre-trained transformer for 3D vision and text alignment," in *Proc. IEEE/CVF Int. Conf. Comput. Vis.*, 2023, pp. 2899–2909.

[20] B. Y. Lin, C. Huang, Q. Liu, W. Gu, S. Sommerer, and X. Ren, "On grounded planning for embodied tasks with language models," in *Proc. AAAI Conf. Artif. Intell.*, vol. 37, no. 11, 2023, pp. 13192–13200.

[21] N. Roy et al., "From machine learning to robotics: Challenges and opportunities for embodied intelligence," in *Proc. 7th Conf. Robot Learn.*, vol. 229, 2023, pp. 2134–2150.

[22] J. Mendez-Mendez, L. P. Kaelbling, and T. Lozano-Pérez, "Embodied lifelong learning for task and motion planning," in *Proc. Conf. Robot Learn.*. PMLR, 2023, pp. 2134–2150.

[23] A. Broering, C. Niedermeier, I. Olaru, U. Schopp, K. Telschig, and M. Villnow, "Toward embodied intelligence: Smart things on the rise," *Computer*, vol. 54, no. 7, pp. 57–68, 2021.

[24] F. Zeng, W. Gan, Y. Wang, N. Liu, and P. S. Yu, "Large language models for robotics: A survey," 2023, *arXiv:2311.07226*.

[25] X. Xiao et al., "Robot learning in the era of foundation models: A survey," 2023, *arXiv:2311.14379*.

[26] A. Kalinowska, P. M. Pilarski, and T. D. Murphey, "Embodied communication: How robots and people communicate through physical interaction," *Annu. Rev. Control, Robot., Auton. Syst.*, vol. 6, pp. 205–232, 2023.

[27] J. Duan, S. Yu, H. L. Tan, H. Zhu, and C. Tan, "A survey of embodied AI: From simulators to research tasks," *IEEE Trans. Emerg. Topics Comput. Intell.*, vol. 6, no. 2, pp. 230–244, Apr. 2022.

[28] Z. Lu and H. Yu, "GTAC-hand: A robotic hand with integrated tactile sensing and extrinsic contact sensing capabilities," *IEEE/ASME Trans. Mechatron.*, vol. 28, no. 5, pp. 2919–2929, Oct. 2023.

[29] J. Wang et al., "Large language models for robotics: Opportunities, challenges, and perspectives," 2024, *arXiv:2401.04334*.

[30] H. Hsiao, J. Sun, H. Zhang, and J. Zhao, "A mechanically intelligent and passive gripper for aerial perching and grasping," *IEEE/ASME Trans. Mechatron.*, vol. 27, no. 6, pp. 5243–5253, Dec. 2022.

[31] P. Gao, P. Wang, F. Gao, F. Wang, and R. Yuan, "Vision-language navigation with embodied intelligence: A survey," 2024, *arXiv:2402.14304*.

[32] D. Gan, S. Caro, G. Chen, S. Nokleby, and J. S. Dai, "Theory and design of reconfigurable mechanisms and robots," *J. Mechanisms Robot.*, vol. 14, no. 6, 2022, Art. no. 060301.

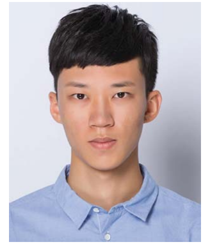
[33] A. Ramesh, P. Dhariwal, A. Nichol, C. Chu, and M. Chen, "Hierarchical text-conditional image generation with CLIP latents," 2022, *arXiv:2204.06125*.

[34] A. Kirillov et al., "Segment anything," in *Proc. IEEE/CVF Int. Conf. Comput. Vis.*, 2023, pp. 3992–4003.

Lei Ren (Senior Member, IEEE) received the Ph.D. degree in computer science from the Institute of Software, Chinese Academy of Sciences, Beijing, China, in 2009.

He is currently a Professor with the School of Automation Science and Electrical Engineering, Beihang University, Beijing. He is also with the Zhongguancun Laboratory, Beijing, and the State Key Laboratory of Intelligent Manufacturing System Technology, Beijing. His research interests include neural networks and deep learning, industrial embodied intelligence, and industrial AI applications.

Dr. Ren is currently an Associate Editor for the IEEE TRANSACTIONS ON NEURAL NETWORKS AND LEARNING SYSTEMS, a Technical Editor for the IEEE/ASME TRANSACTIONS ON MECHATRONICS and other international journals.



Jiabao Dong (Graduate Student Member, IEEE) received the B.S. degree in automation science from the School of Automation Science and Electrical Engineering, Beihang University, Beijing, China, in 2020, where he is currently working toward the Ph.D. degree in pattern recognition and intelligent systems with the School of Automation Science and Electrical Engineering.

His research interests include industrial artificial intelligence, embodied intelligence, and robotic manipulation.

Shuai Liu (Student Member, IEEE) received the B.S. degree in intelligent manufacturing engineering from the School of Mechanical Engineering, Shandong University, Jinan, China, in 2023. He is currently working toward the master's degree in control science and engineering with the School of Automation Science and Electrical Engineering, Beihang University, Beijing, China.

His research interests include robotic manipulation and grasp detection.

Lin Zhang (Senior Member, IEEE) received the B.S. degree in control science from Nankai University, Tianjin, China, in 1986 and the M.S. and Ph.D. degrees in control science from Tsinghua University, Beijing, China, in 1989 and 1992, respectively.

He is currently a Professor with Beihang University, Beijing. He was the Chair of the Society for Modeling and Simulation International (SCS) and is currently the Chair of Asian Simulation Conference (ASIASIM). His research interests include modeling and simulation, cloud manufacturing, and model engineering.

Dr. Zhang is the Fellow of SCS, ASIASIM, Executive Vice President of China Simulation Federation, and an Associate Editor-in-Chief, Associate Editors for six peer-reviewed international journals.

Lihui Wang received the Ph.D. degree in intelligence science from Kobe University, Kobe, Japan, in 1993.

He is currently a Chair Professor with the KTH Royal Institute of Technology, Stockholm, Sweden. He has authored or coauthored ten books and authored more than 700 scientific publications. His research interests include cyber-physical systems, real-time monitoring and control, human–robot collaboration, and adaptive manufacturing systems.

Dr. Wang is the Editor-in-Chief of *International Journal of Manufacturing Research*, *Journal of Manufacturing Systems*, and *Robotics and Computer-Integrated Manufacturing*. He is a Fellow of Canadian Academy of Engineering (CAE), Fellow of International Academy for Production Engineering (CIRP), Fellow of Society of Manufacturing Engineers (SME), and Fellow of American Society of Mechanical Engineers (ASME). He is also a Registered Professional Engineer in Canada, the President of North American Manufacturing Research Institution of SME from 2020 to 2021, and the Chairman of Swedish Production Academy from 2018 to 2020.