DOI 10.15302/J-SSCAE-2026.04.028

工业具身智能应用场景与技术发展研究

彭乐乐，李秀萍，郗欣甫，陈玉洁，孙以泽 *

（东华大学机械工程学院，上海 201620）

摘要：智能制造水平是衡量工业制造核心竞争力的关键标志，具身智能技术为工业制造系统实现战略升级提供了重要支撑，推动工业实体从被动执行向主动感知、主动适应与主动交互转型。区别于底层算法原理的传统研究，本文以工程落地为研究定位，针对现有基础模型难以直接跨越工业场景中“高精度、快节拍、强鲁棒”工程鸿沟的问题，提出了“场景痛点反推技术需求”的分析框架。我国智能制造正加速向柔性化、自主化、融合化方向转型，工业场景复杂工况、严苛工艺指标对智能技术落地提出更高要求，通用智能模型与工业实际需求脱节的问题愈发突出。为此系统性梳理了工业具身智能在人机协作、精密装配、智能运维、物流仓储及数字孪生五大典型工业应用场景中的技术赋能，构建了包含多模态感知与状态特征、认知推理与任务决策、灵巧执行与运动控制、数字孪生与闭环优化的四层级技术体系，提炼了具身智能的工程化适配与演进路径，以实现场景与技术的双向匹配。为推动工业具身智能技术的发展，研究建议，加强技术攻关，突破技术瓶颈；完善产业生态，加快工业化落地；强化政策支持，优化发展环境，实现具身智能从实验验证走向工业场景可控、规模化部署，探索智能制造落地新模式。

关键词：多模态感知；人机协作；工业具身智能；工业场景应用；智能制造

中图分类号：TP18；TP24　　**文献标识码：**A

Application Scenarios and Technological Development of Industrial Embodied Intelligence

Peng Lele, Li Xiuping, Chi Xinfu, Chen Yujie, Sun Yize *

(College of Mechanical Engineering, Donghua University, Shanghai 201620, China)

Abstract: The advancement of intelligent manufacturing is a critical benchmark for the core competitiveness of the industrial sector. Embodied intelligence provides crucial support for the strategic upgrading of manufacturing systems, driving a paradigm shift in industrial entities from passive execution to active perception, adaptation, and interaction. Distinct from traditional reviews focused on underlying algorithmic principles, this study targets engineering implementation. To address the inherent challenge that current foundation models struggle to bridge the engineering gap of high precision, rapid cycle times, and strong robustness in industrial settings, we propose an analytical framework that derives technological requirements by reverse-engineering scenario-specific pain points. Intelligent manufacturing of China is currently advancing rapidly toward flexibility, autonomy, and integration. The complex working conditions and stringent process indicators in industrial scenarios have raised higher requirements for the implementation of intelligent technologies, and the disconnection between general intelligent models and practical industrial demands has become increasingly prominent. Therefore, this study examines the technological empowerment of industrial embodied intelligence across five representative industrial application domains: human–robot collaboration, precision assembly, intelligent operation and maintenance,

收稿日期：2026-04-13；**修回日期：**2026-06-12

通讯作者：*孙以泽，东华大学机械工程学院教授，中国工程院院士，研究方向为具身智能与智能制造；E-mail: sunyz@dhu.edu.cn

资助项目：中国工程院咨询项目“福建省运动鞋服功能纺织材料及产品创新发展研究”(2025-DFZD-06)

本刊网址：sscae.engineering.org.cn

logistics and warehousing, and digital twin. A four-tier technical architecture is established, comprising multimodal perception and state characterization, cognitive reasoning and task decision-making, dexterous execution and motion control, and digital twin-enabled closed-loop optimization. Furthermore, the engineering adaptation and evolutionary pathways of embodied intelligence are delineated, thereby achieving bidirectional alignment between application scenarios and technological capabilities. To advance the development of industrial embodied intelligence, this study recommends the following strategic directions: strengthening frontier research to overcome critical technical bottlenecks, cultivating a robust industrial ecosystem to accelerate large-scale deployment, and enhancing policy support to optimize the developmental landscape. These efforts are expected to facilitate the transition of embodied intelligence from laboratory validation to controllable and scalable industrial implementation, thereby exploring novel paradigms for intelligent manufacturing.

Keywords: multimodal perception; human–machine collaboration; industrial embodied intelligence; industrial scenario application; intelligent manufacturing

一、前言

目前，我国工业产值约占全球比重的30%，智能制造水平与数字化转型能力已成为衡量工业制造竞争力的重要标志[1]。我国在新能源汽车、航空航天装备、高端制造装备、轨道交通及智能驾驶等领域已建成一批智能化工业产业园区，工业制造正从传统规模化生产模式向柔性化、自主化、融合化方向纵深演进[2-4]。然而，当前工业制造在柔性生产适配能力、复杂环境设备安全维护、全流程数据协同与决策联动能力等方面仍然存在不足，主要表现为：在生产制造环节，面临从机器主导的单一化生产向人机协同交互的新型生产模式转变；在装备安全监测方面，面对复杂环境和突发状况，亟需实现从被动预警辅助向主动处置执行的功能升级；在生产流程管控领域，需完成从单向数据观察到虚实融合、双向赋能的范式革新[5-6]，其本质变化在于工业系统由被动执行向主动感知、主动适应与主动执行转变[7]。视觉–语言模型、机器人基础模型、多模态传感与高性能控制等技术的发展，使智能体具备了在更复杂环境中进行任务理解、动作生成与交互执行的可能性，具身智能由此成为制造领域的重要研究前沿[8-11]。具身智能技术的发展和应用将助力实现工业制造技术的战略升级。

在工业场景中，具身智能更加关注智能体在真实工业环境中的行动能力，即通过物理载体与工件、设备、环境以及人员形成持续耦合，在工艺约束和安全规范下完成感知、认知、决策与执行闭环[7,12]，其不仅能替代人工完成高危、重复、精密作业，更能成为人机共融的生产伙伴，重塑作业流程。具体覆盖五大核心场景：在人机协作中通过“感知–决策–执行”闭环实现协同作业，在柔性生产中实现柔性加工与精密装配，在智能运维中完成从监测到一体化执行、物流仓储中破解非结构化环境的作业难题，在数字孪生中实现高精度实时虚实双向赋能。这决定了工业具身智能不是将通用大模型简单部署到机器人系统中，而是面向工业任务将算法模型、执行机构、感知系统、控制架构、知识规则和工程约束进行系统集成的综合智能形态[13-15]，体现了其从被动辅助到主动交互的本质升级。

现有研究多聚焦于基础理论、基础模型和世界模型的构建[16-17]，探索大模型对任务理解与执行能力的提升，但其核心面向开放世界场景，未充分考虑工业场景的强约束、高可靠、高精度要求，难以直接落地。部分研究对工业机器人、数字孪生、人机协作与自主运维等方向开展了关键技术综述和进展梳理[18-25]，但缺乏将具身智能作为核心体系，开展跨场景、全链条的系统研究，未形成场景与技术的反向适配逻辑。我国发布的相关数字经济与智能制造规划的政策文件也指出，大力推动新型交互式智能技术赋能实体工业制造。尽管研究取得进展，但当前工作缺乏面向工业落地的实施路径，忽视了工程环境约束与系统集成的落地端研究。为此，本文区别于侧重基础理论的研究，采用从场景驱动到技术反推的思路，旨在回答“具身智能如何适配工业现场工程约束（如微米级装配精度、毫米级响应节拍、刚性碰撞安装底线）”这一核心问题，弥补现阶段理论大模型与工业严苛工程指标之间的研究空白。通过构建工业具身智能的技术体系架构，剖析关键技术突破点与现存瓶颈，展望未来发展趋势，推动具身智能由实验室技术验证逐步走向工业场景部署，探索在工业制造中具身智能技术实施新途径。

二、工业具身智能的概念、内涵与演进过程

（一）工业具身智能的概念、内涵

1. 概念及内涵

传统的离身智能依托大模型在虚拟环境中进行感知、识别与预测，具备一定的信息处理与符号推理能力，但缺乏对真实物理环境的直接干预与适应。然而，具身智能强调智能并非孤立存在于抽象计算系统中，依托物理载体，通过与环境的持续交互在感知、行动、反馈中形成和修正[26]。将这一思想引入工业制造场景后，工业具身智能可定义为：智能体依托工业物理载体，在制造环境中通过多模态感知[27]、任务理解[28]、自主决策与动作执行形成闭环交互能力[29-30]，从而在工艺约束和安全约束下完成复杂工业任务的一类智能形态。这一定义包含四层内涵：一是工业具身智能以机器人、移动平台、复合装备、工站或其他可执行实体为依托，不能脱离物理实体；二是工业具身智能的目标不是简单地生成信息，而是面向抓取、搬运、装配、巡检、维护、检测等工业任务实现可验证的行动结果；三是工业具身智能的智能形成依赖于真实工业环境中的交互，包括与设备、工件、工具、工艺流程及人员的耦合；四是工业具身智能的运行受到制造精度、节拍、安全、质量一致性和可追溯性等工程条件的严格约束。因此，工业具身智能的研究对象既不同于传统自动化中的单一执行机构，也不同于只具备知识推理能力的通用大模型，而是将感知、控制、学习、认知与工程实现统一起来的复杂系统。

工业具身智能并非现有智能制造技术的简单统称，更不能将所有数字化、自动化技术均纳入其范畴。相较于依赖预定义逻辑、固定工位和明确边界的传统工业自动化[4]，工业具身智能更强调具有在来料波动、工件公差、环境扰动和人员介入等不确定条件下完成任务的能力，适用于传统刚性自动化难以经济覆盖的半结构化场景。与主要用于分析、预测和优化的工业人工智能不同[31]，工业具身智能不仅要“理解”“判断”，还要能够作用于物理世界，将感知结果转化为动作决策与执行行为，具有更强的行动导向。与通用具身智能或服务机器人相比，工业场景对精度、节拍、安全、可追溯性和责任边界要求更高，不接受“近似完成”。因此，工业具身智能既不是单纯的硬件载体（如协作机器人、人形机器人），也不是纯粹的赛博空间算法（如工业人工智能、工业智能体、数字孪生），而是在工业物理与工艺约束下，具备自主感知、认知决策与闭环物理执行能力的综合智能形态，其核心不只是智能性，更在于工艺知识约束下的可靠控制与工程可部署性。

2. 特征

工业具身智能是在智能制造体系下，依托具身智能技术实现感知、认知、决策与执行深度耦合的新型工业智能形态，以工业物理系统为依托，以制造任务为核心对象，强调智能体在真实环境中的自主感知、实时决策与物理交互能力，旨在提升制造系统对复杂任务、动态扰动和非结构化场景的适应能力。工业具身智能的特征主要有：一是具身性，即智能能力嵌入具体工业载体之中，并受到本体结构、执行机构和传感系统等条件约束[32]；二是任务导向性，以装配、搬运、检测和维护等工业任务的有效完成为目标；三是闭环性，在感知、理解、决策、执行与反馈之间形成实时循环；四是工艺约束性，即行为生成必须服从精度、质量、节拍与安全等工业要求；五是可控性，系统需具备可验证、可干预、可审计和可维护等工程属性。

3. 本质属性

从本质上看，工业具身智能是一种面向工业物理世界、以任务执行为中心的系统智能。它不是单一算法或单一装备的能力延伸，而是在“本体–环境–任务”耦合关系中形成的综合智能形态。工业具身智能的能力来源不仅包括数据驱动学习，还包括工艺知识、控制理论、物理规律和场景经验的共同支撑。因此，工业具身智能既具有人工智能系统的学习与泛化属性，又具有工业系统所要求的稳定性、可控性和可部署性。其意义不在于替代传统自动化，而在于弥补传统自动化在复杂任务、动态适应和非结构化场景中的不足，推动制造系统向高柔性、高韧性和高自主性方向演进。

（二）工业具身智能技术演进

工业生产方式的演进始终伴随着技术进步与组织模式变革。当前，智能制造已由以自动化、信息化为主的发展阶段，逐步迈向以数字化、网络化和智能化深度融合的新阶段。然而，传统基于“感

知–通信–控制”架构的生产系统，在复杂、动态、非结构化工业场景中仍存在明显局限：一方面，其运行逻辑高度依赖预设程序和固定轨迹，对环境适应性不足；另一方面，感知、决策与执行之间仍存在一定程度的分离，对接触状态、力学约束及动态交互过程的理解和利用能力有限。具身智能的引入，推动工业智能系统由以离线计算和规则执行为主的智能形态向面向真实物理交互的在线智能演进。其核心在于，智能体依托物理载体，在真实环境中通过持续感知、行动与反馈闭环形成对任务、环境及物理规律的理解，从而实现感知、认知与执行的一体化，构建面向复杂工业场景的世界模型。这一过程不仅提升了设备和系统的自主能力，也推动了生产模式、运维模式和管控模式的整体变革，主要体现在以下3个方面（见图1）。

1. 交互式生产模式的革新

传统刚性自动化生产线以大规模、标准化生产为主要目标，其系统设计强调工艺稳定性和流程确定性，但对任务变化、环境波动及人机协同需求的适应性不足，因而往往面临换产成本高、调试周期长、柔性不足等问题，难以满足多品种、小批量和个性化制造的发展需求[6]。具身智能的发展推动生产模式由预设执行向交互适应转变，使制造系统具备面向任务、环境和人员变化的动态响应能力。依托视觉、力觉、触觉等多模态感知及学习控制方法，具身智能体能够在感知环境状态的基础上实时调整动作策略，并通过迁移学习、模仿学习和强化学习提高对新任务与新场景的适应能力。同时，大模型增强了工业系统对自然语言指令和复杂任务语义的理解能力，降低了传统机器人对人工编程和示教的依赖，推动人机关系由“指令执行”向“人机协同决策与任务共担”的交互式生产模式演进。例如，已具备机械装配经验的智能体，可在电子元件组装、零部件抓取等任务中迁移已有操作知识，并结合实时反馈完成动作优化。

2. 自主式设备安全监测升级

设备安全监测是工业系统稳定运行的重要保障。传统设备安全监测主要依赖固定传感器与预定义规则，能够实现基本状态识别和故障报警，但在复杂工况和隐蔽故障场景下，仍存在感知维度有限、诊断能力不足和响应滞后等问题。具身智能推动设备监测由被动感知、事后处置向主动感知、自主响应转变。通过融合视觉、热成像、振动、声学等多模态信息，具身智能体能够更全面地获取设备运行状态，增强对异常征兆和潜在风险的早期识别能力。在此基础上，结合数据驱动与机理约束方法，系统可进一步实现对故障特征、演化趋势及风险传播的综合分析。更重要的是，具备移动与操作能力的智能体能够在识别异常后实施自主复核和现场处置，形成“感知–诊断–决策–执行”的闭环监测体系，提高工业系统的主动防御能力。

3. 融合式全流程管控范式转变

现有工业全流程管控虽已实现计划、调度、执行与反馈环节的初步联通，但不同系统和执行单元之间仍普遍存在信息分散、协同不足和响应滞后的问题，制约了制造系统的实时优化与柔性重构能

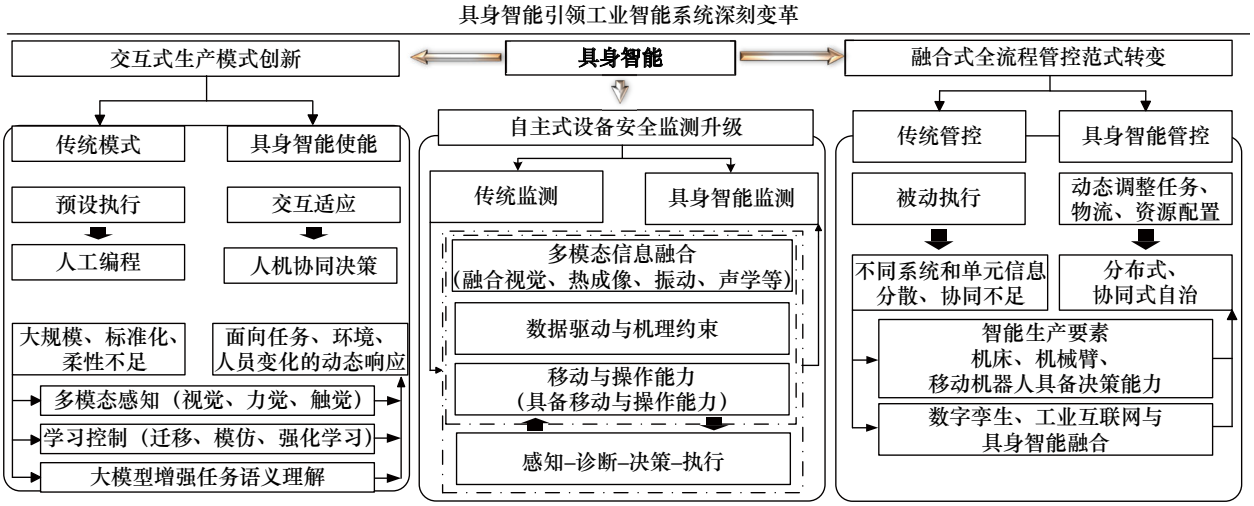


图1　工业具身智能模式变革

力。具身智能的发展推动制造管控模式由集中式、层级式控制向分布式、协同式自治演进。随着机床、机械臂、移动机器人和物料单元等生产要素逐步具备感知、决策和通信能力，其角色由被动执行单元转变为可参与局部决策与协同优化的智能主体。与此同时，数字孪生、工业互联网与具身智能的融合，强化了物理空间与数字空间之间的实时映射和闭环联动，使系统能够根据订单变化、设备状态和现场扰动，动态调整任务分配、物流路径和资源配置。由此，制造系统逐步形成局部自治与全局优化相融合的运行机制，为构建兼具柔性、自主性、韧性与效率的新型生产体系提供支撑。

三、工业具身智能典型应用场景

具身智能技术在工业领域的应用正从简单的替代人力向复杂的人机协作、从单一工序向全流程优化赋能演进，其在提升生产效率、降低运营成本、增强系统柔性等方面的价值已得到初步验证。针对人机协作、柔性生产、智能运维、数字孪生和物流仓储的典型工业场景，分析具身智能技术的应用模式与效果，以获取其应用框架（见图2）。

（一）人机协作

在工业5.0“以人为本”的制造理念下，人机协作虽能有效改善作业环境，但其深层矛盾日益凸显。一方面，在搬运重物、持续姿态作业及危险区域操作等场景中，高强度、重复性和高风险的作业环节给人员带来了极大的生理负荷与安全隐患。另一方面，现有机器人高度依赖基于物理隔离或系统限流的被动安全机制，制约了共享空间内的作业柔性与空间利用率。此外，传统机器人基于预编程的程序化执行，难以适应任务节奏、工件状态和人员行为的强随机性，导致生产系统在复杂工况下的整体运行效能难以达到预期。为克服上述痛点，人机协作系统亟需在认知与协同维度实现深度的能力升级，建立“人判断、机执行”的优势互补机制，以充分融合人类的环境适应、经验判断、异常处理能力与机器人在重复执行、节拍稳定、持续作业方面的专长。同时，机器人需具备语义驱动的自然交互能力，实现从代码驱动到语义驱动的跨越。面对高动态作业环境，人机协作系统需具备实时自适应决策能力，从而自主重构动作轨迹与任务分工；在安全维度上，需完成从被动物理防护向基于环境认知与行为预测的主动前瞻性安全响应的范式转换。

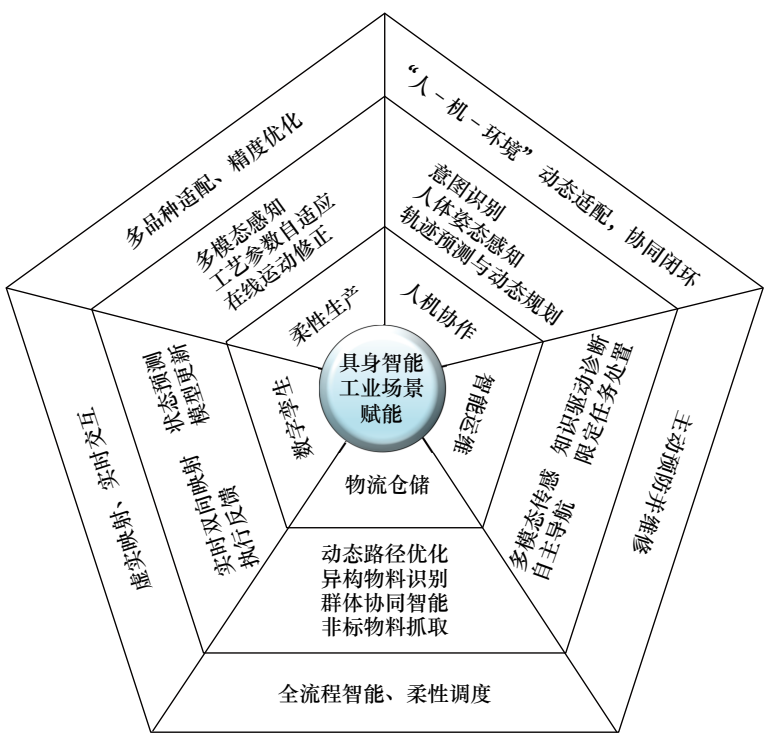


图2 具身智能在工业应用场景的赋能

具身智能的引入为上述能力构建提供了核心技术框架。首先，基于多源异构信息（如视觉、力觉、触觉及语音等）的深度融合感知技术，为复杂协作构筑了高维状态表征。其次，依托强大的语义建模与跨模态对齐技术，打通了自然语言指令、物理空间映射与底层运动控制的壁垒。再次，基于持续感知与策略闭环的动态控制架构，赋予了系统在不确定环境下的鲁棒性。此外，融合示教学习、强化学习与模仿学习的技能迁移算法，使机器人能够习得精密装配、柔性抓取等强经验属性的接触操作，并在交互反馈中实现技能泛化与精度迭代。例如，在蔚来汽车总装车间中，优必选 Walker S 人形机器人与人类产线工人协同完成车门锁质检及安全带检测。相较于传统工业机器人的固定隔离作业，引入具身大模型的人机协作系统将意图识别延迟降至 100 ms 以内。在保证《机器人与机器人装备 协作机器人》（ISO/TS 15066：2016）中提及的人机协作安全标准的前提下，共享工位的整体作业节拍缩短了 20%，漏检率下降至 0.1% 以下。

具身智能驱动的人机协作展现出巨大潜力，但工程化落地仍面临多重技术壁垒。在复杂且伴有强干扰的工业现场中，机器对人类多模态意图的精准捕捉与消歧仍存在较高误判率；面对非结构化的动态共享物理空间，实现兼顾低延迟与高可靠性的实时安全风险量化评估仍具严峻的工程挑战；在涉及多维精细力控与高频交互的复杂接触任务中，机器人的柔顺控制与泛化适应能力尚未达到工业级替代要求。上述瓶颈制约了人机协作系统由被动跟随向主动认知协同的全面升维，具身智能赋能人机协作路径如图 3 所示。

（二）柔性生产与精密装配

柔性生产与精密装配作为高端制造的核心场景，需具备多品种快速换型与微小公差高质量制造能力。然而，在复杂工况下，现有制造体系正面临多维度的技术痛点。一是高精度感知能力显著不足。在孔轴配合、异形件抓取、复杂曲面打磨及精密涂胶等任务中，由于微小特征难以提取、工件相互遮挡、表面反光及姿态不确定等物理限制，单一感知手段极易失效，导致复杂接触状态难以准确表征。二是接触密集型任务的底层控制难度极高，如插入、压装、拧紧与卡扣配合等作业，对机器人的高精度轨迹控制、力位协同调节及在线误差补偿提出了严苛的约束要求。此外，传统系统高度依赖专用工装夹具、繁琐的离线编程与固有的人工经验，在面对频繁的新品换型与工艺变更时，现有设备面临重构周期长、成本高且泛化能力弱的问题，难以满足现代高端制造对敏捷响应的迫切诉求，如在新能源汽车底盘管线装配、中国强制性产品认证中的电子异形件插接或运动鞋服功能面料缝制中，传统系统仍依赖刚性工装。面对来料公差（如±0.5 mm 漂移）、反光金属表面造成的视觉盲区以及非标零件换型，现有设备重新编程周期长，微小特征装配极易因累计误差导致卡滞或压伤。

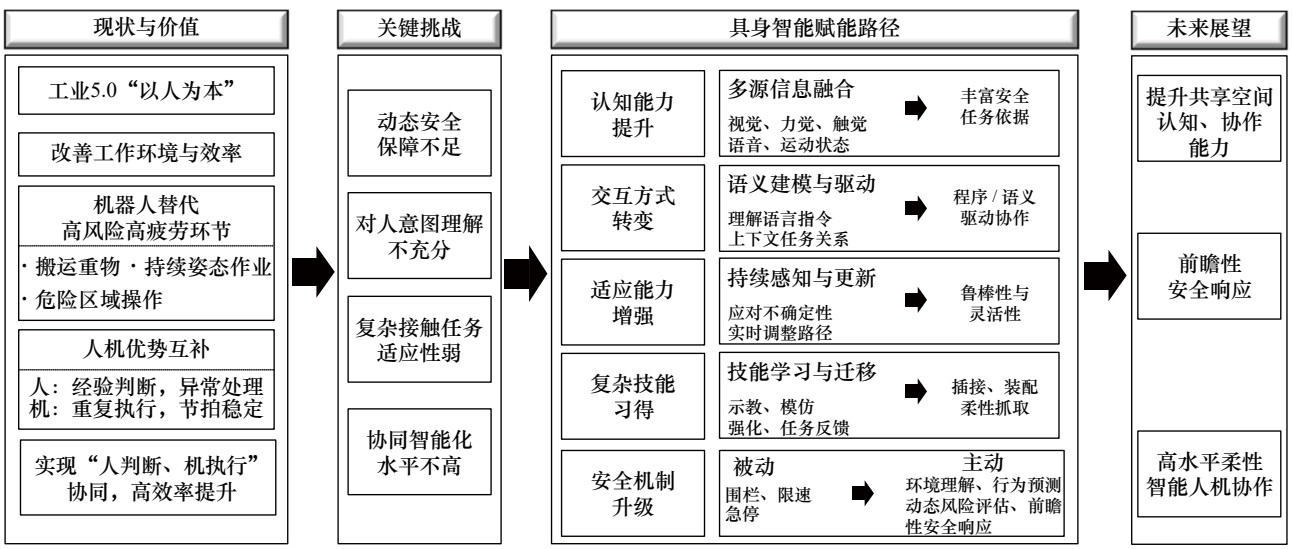


图 3　具身智能赋能人机协作

为破解上述制造困境，需要系统具备接触状态的动态感知能力与亚毫米级的位姿在线纠偏能力。智能体需要通过“试探–感知阻力–微调姿态–发力”完成操作。这意味着机器人需要在高度动态的接触过程中，自主完成精准对位、插入力调节与姿态微调，有效吸收工艺偏差累积与外部环境扰动。通过这种高度的自适应能力，生产系统才能在提升多样化任务响应速度的同时，确保极高的装配质量与制造一致性。

具身智能技术的融入为重塑复杂柔性装配闭环提供了核心突破路径。引入基于多源信息融合（视觉+六维力矩）的具身推理模型，在视觉提供全局粗定位后，切换至基于强化学习的柔顺阻抗控制模型。通过在线辨识接触界面的刚度矩阵，动态调整插入力阈值（如保持在5N~10N），实现卡扣配合或孔轴微调，从而大幅提升精密操作的成功率、接触系统稳定性及表面加工一致性，降低因偏差累积或工况扰动带来的质量风险。以上海非夕机器人科技有限公司的自适应机器人在笔记本电脑外壳精密打磨及内存条柔性插接中的应用为例，传统离线编程的机械臂在曲面打磨时受来料公差影响，良率通常在90%左右波动；引入具备“手眼协调”的具身自适应机器人后，系统可实现0.1N级别的力控精度与±0.05 mm的轨迹公差。对比传统方案，其换型调试时间从1周缩短至4 h，柔性装配与打磨的一次交验合格率提升至99.9%。

尽管具身智能展现出显著的技术优势，但其向高端柔性制造场景的规模化落地仍受限于多重瓶颈。当前的最大阻碍在于高频控制的计算延迟与模型的黑盒特性。深度学习策略的推理延迟（通常>50 ms）难以匹配工业精密装配底层控制环（<2 ms）的需求，且基于神经网络的动作生成缺乏可解释性，一旦发生越界发力，现有的具身架构难以及时在毫秒级触发机械硬件的断电保护机制。在面对超精细公差或极端物理特性的精密制造任务时，多模态模型的泛化表征精度与底层控制执行的实时鲁棒性往往难以达到工业级“绝对可靠”的严苛标准。此外，柔性生产中长尾异常工况的不可预见性，使机器人在脱离特定人工示教和结构化约束后，面向未知工艺变更的自主泛化适应性仍有待进一步跨越（见图4）。

（三）智能运维

智能运维以设备全生命周期健康管理为核心，依托工业物联网与边缘计算，初步实现了设备状态的连续感知。然而，面向高危、受限空间及复杂环境，现有运维范式在实际工业场景中仍面临挑战[33]。一是工业现场多源传感数据常伴随强噪声、高缺失率及标准不一等问题，加之真实故障样本稀缺，导致传统数据驱动模型在跨设备应用时泛化能力严重受限。二是面对复杂多变的工况扰动与设备异构性，现有诊断预测模型稳定性不足且缺乏可解释性，决策过程的“黑盒化”难以获取现场人员的信任。更为关键的是，当前多数运维系统受限于“监而不控”的开环阶段，缺乏从异常识别到物理干预的联动能力，导致高危及狭小空间内的局部维护仍高度依赖人工，存在极大的安全隐患与停机风险。

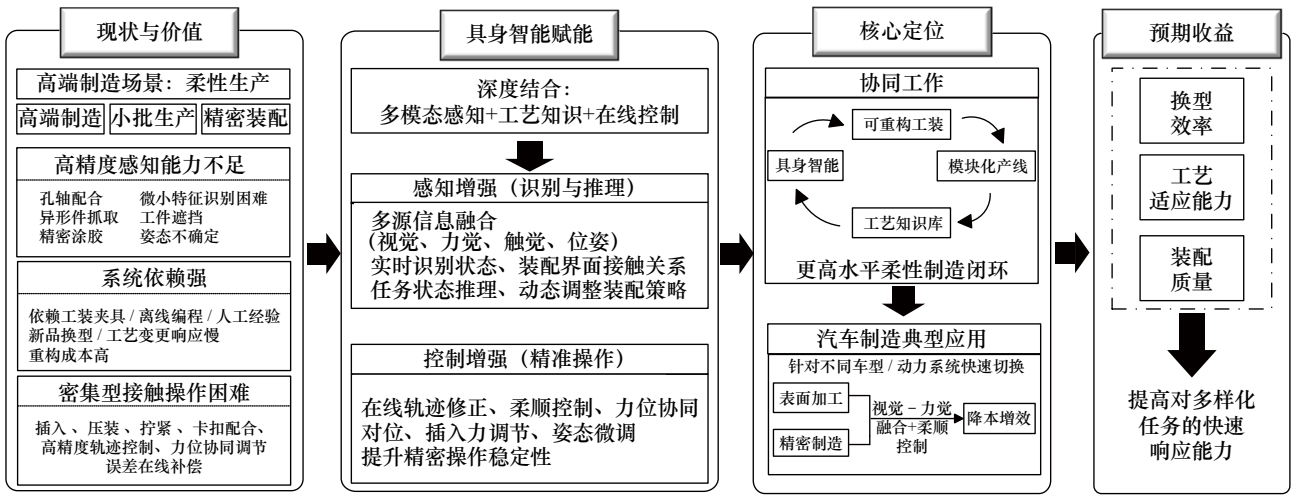


图4　具身智能赋能柔性生产与精密装配

针对上述挑战，新一代工业运维系统的演进亟需突破单一的信息域框架，向物理域深度延伸，并由此反向推导出三大核心能力需求：全域多模态的强鲁棒感知、可解释的自主认知推理以及复杂受限空间内的灵巧物理交互。具身智能满足了上述需求，是重塑运维范式的关键赋能技术。具身智能将移动平台、多模态传感、知识驱动认知与自主控制深度融合。在感知层面，通过整合视觉、热成像、声振等多维物理量克服单一传感器的脆弱性；在决策层面，结合世界模型与先验机理知识，赋予系统对物理世界的因果推理能力，大幅提升跨设备泛化性与诊断可解释性；在执行层面，依托蛇形机械臂、轨道巡检机器人等特种平台，结合精准力控与在位操作算法，替代人类深入危险区域执行结构检测、样本采集与局部维修。这一技术群的融合，推动工业运维的技术演进路线从初级的单模态状态监测、基于大模型与数字孪生的虚拟认知诊断，最终迈向以具身智能体为物理代理的“感知–推理–执行”全自主闭环，实现从单纯的“发现问题”到自主“解决问题”的跃升，将运维模式由被动响应转变为主动预防与干预。例如，中控技术股份有限公司联合使用宇树科技股份有限公司研发的四足机器人在氯碱化工电解槽车间中进行危化巡检，实现了对近5000根碱液管的24 h无人化覆盖。对比人工与传统挂轨机器人，其复杂故障的漏检率由15%降至1%以内，误报率控制在3%以下，单次全面巡检耗时缩减了80%。

智能运维迈向规模化部署需跨越一系列严峻的工程落地瓶颈。在跨域集成方面，具身智能体亟需与传统监控与数据采集系统（SCADA）、分布式控制系统（DCS）及企业资源计划（ERP）等工业系统进行深度的数据流转和接口互通，面临较高的协议壁垒。在硬件可靠性方面，面对强电磁干扰、高频振动及极端温 / 湿度等恶劣工况，具身传感器的精度、机械机构的寿命以及端侧算力平台的稳定性仍面临极限考验。此外，工业场景对任务协同的实时性要求极高，“云边端”架构下的通信延迟、大规模具身智能体集群的调度算法以及高昂的硬件定制与部署成本，依然是制约具身智能在工业运维领域全面推广的核心挑战，具身智能赋能智能运维路径（见图5）。

（四）物流仓储

物流仓储作为连接生产与配送的关键枢纽，尽管已具备较好的数字化与自动化基础，但在向高度智能化演进的过程中仍面临显著的痛点。具体而言，仓内货物包装形态多变且常伴随堆叠、遮挡与反光，导致传统设备在目标识别与位姿估计上的稳定性严重不足；同时，真实作业环境往往具有非结构化特征，使仅具备定点导航能力的传统移动机器人，如自动导引车（AGV）和自主移动机器人（AMR），难以应对异形件、软包装与混装货物的柔性操作；此外，大规模人机混合作业中的多机路径规划与动态避障极易引发系统性拥堵与安全隐患。

基于上述工业痛点，现代仓储系统亟需反向推

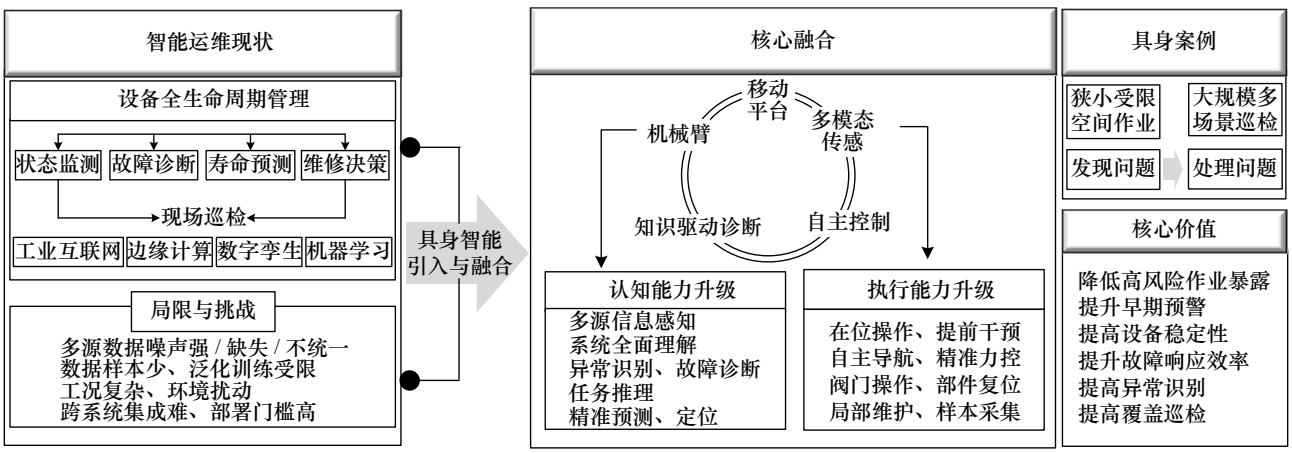


图5　具身智能赋能智能运维

导并重构其底层技术能力。在感知层，需要能够应对复杂干扰与非标准物料的鲁棒性多模态三维（3D）环境理解能力；在决策层，要求系统具备基于实时订单、库存与工位负载的全局动态任务编排及多智能体协同调度能力；在执行层，亟需打破单一“承载搬运”的限制，赋予机器人针对异构物料的柔性抓取与在位装箱能力。具身智能技术的引入，将多模态视觉、移动底盘与机械臂复合平台、大模型及数字孪生有机融合，推动仓储物流由“传统单机定点自动化搬运”向“识别、拣选、转运、配送一体化的复合操作”跃升，并最终演进为具备高环境自适应与群体协同能力的完全自主决策系统（见图 6）。以极智嘉复合移动机器人（AMR+机械臂）在一汽丰田华南零件中心的数智化升级项目为例，面对万级库存量单位（SKU）且形态各异的汽车零部件，传统点到点 AMR 仅能完成搬运，无法解决末端理货难题。在引入具身复合机器人后，通过大模型任务拆解，实现了端到端的自主拣选。相关数据表明，空间利用率提升了 40%，订单响应及拣选效率提升了 20%，对异形复杂件的抓取准确率达到了 99.99% 的工程极限。

物流仓储的大规模工程落地仍存在多项现实瓶颈，如具身复合机器人在高动态人机混合场景下的绝对安全协同与合规性难题、海量机器人在全局动态调度下面临的毫秒级通信延迟与算力负荷以及高昂的柔性执行器定制和跨系统集成成本，这些工程门槛是当前智能仓储真正实现柔性与效率双轨并跃所需跨越的障碍。

（五）数字孪生

数字孪生技术通过物理对象的实时映射与仿真预测，显著提升了制造系统运行的可视化水平与资源配置效率。但在面对复杂工业场景时仍存在突出的现实痛点：多设备耦合、多物理场交互与动态工况的交织，使得高保真建模与虚实强同步举步维艰；异构系统间数据格式与通信接口的标准缺失，严重制约了跨层级互操作与规模化集成；更核心的是，当前多数系统仍停留在状态监测与仿真分析的开环阶段，虚拟空间产生的决策难以向物理执行空间进行有效下达与联动。

针对上述痛点，数字孪生系统亟需反向推导并重构其底层技术能力：在感知层，需要能够精准捕捉多物理场耦合与环境动态变化的高保真多维映射技术；在决策层，要求系统具备融合异构数据进行实时推理与前瞻性调度优化的认知技术；在执行层，需要打破虚实隔离，具备从虚拟空间决策到物理空间干预的精准闭环控制能力。具身智能将多模态感知、自主决策与物理执行有机结合，使数字孪

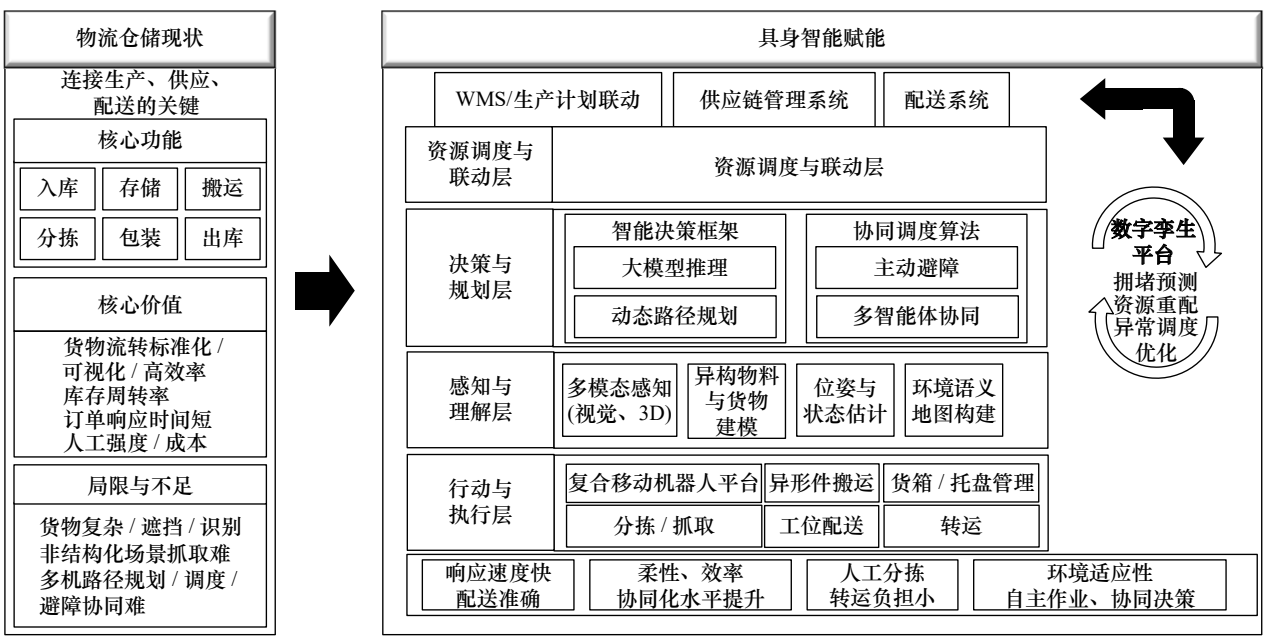


图 6　具身智能赋能物流仓储

生由单纯的“映射与分析”拓展为“感知–决策–执行”一体化系统，深刻还原了真实工况下人、机、物之间的复杂交互关系，实现场景赋能。这一技术的融合推动工业系统由传统静态模型表征向虚实动态映射与仿真预测过渡，并最终迈向具身自主决策与物理闭环联动的完全虚实协同范式（见图7）。宝马全球工厂引入NVIDIA Omniverse与Isaac Sim平台，构建了工业元宇宙具身智能训练基地，利用高保真孪生底座，机器人的路径规划与柔顺装配策略在云端并行训练后直接下发，相比传统物理调试，产线规划时间缩减了30%，新车型换产的机器人现场部署与调试周期由数周大幅压缩至数天（效率提升超70%），极大降低了物理试错成本。

数字孪生虚实协同范式的工程落地实现，面临一系列严峻的瓶颈制约，包括高频海量数据在采集、传输与计算过程中的极低延迟和高扩展性挑战，异构软硬件体系集成时的底层互操作壁垒，高度动态的虚实双向交互中保障工业级绝对安全的系统鲁棒性难题，制约了数字孪生的跨层级集成与规模化部署。

综上所述，工业具身智能在人机协作、柔性装配、物流仓储、智能运维及数字孪生等场景中的应用，并非单纯的算法平移，而是深度嵌套于各类工业垂直领域的工艺与物理约束之中。不同场景对智能体的感知模态、控制精度、安全底线及评测指标有不同的工程要求。为了清晰界定具身智能在各场景下的应用边界与技术差异，本研究从核心痛点、动作空间、控制要求及落地瓶颈等工程维度，对上述五大典型应用场景进行了系统性梳理与横向对比（见表1）。

四、工业具身智能关键技术体系

工业具身智能的关键技术体系是涵盖“感知–决策–执行–优化”的动态闭环反馈系统，主要由多模态感知与状态表征、认知推理与任务决策、灵巧执行与运动控制、数字孪生与闭环优化四大核心模块构成（见图8）。多模态感知与状态表征通过视觉、力觉、触觉等多源信息的高效融合，提升智能体对工件状态、接触关系和环境变化的识别能力，为系统在非结构化场景下的稳定作业筑牢数据基底。认知推理与任务决策负责将多维感知结果转化为全局任务规划和局部动作策略，使系统能够在严苛的工艺约束、安全边界和时变工况下自主完成任务拆解与在线自适应调整。灵巧执行与运动控制决定了智能体能否将“会判断”的策略有效转化为“会完成”的物理操作，其核心在于跨越软/硬件鸿沟，实现高精度、强鲁棒与安全可控的力位协同。数字孪生与闭环优化依托高保真虚实映射，打通策略训练、虚拟验证、在线运行与持续迭代的全

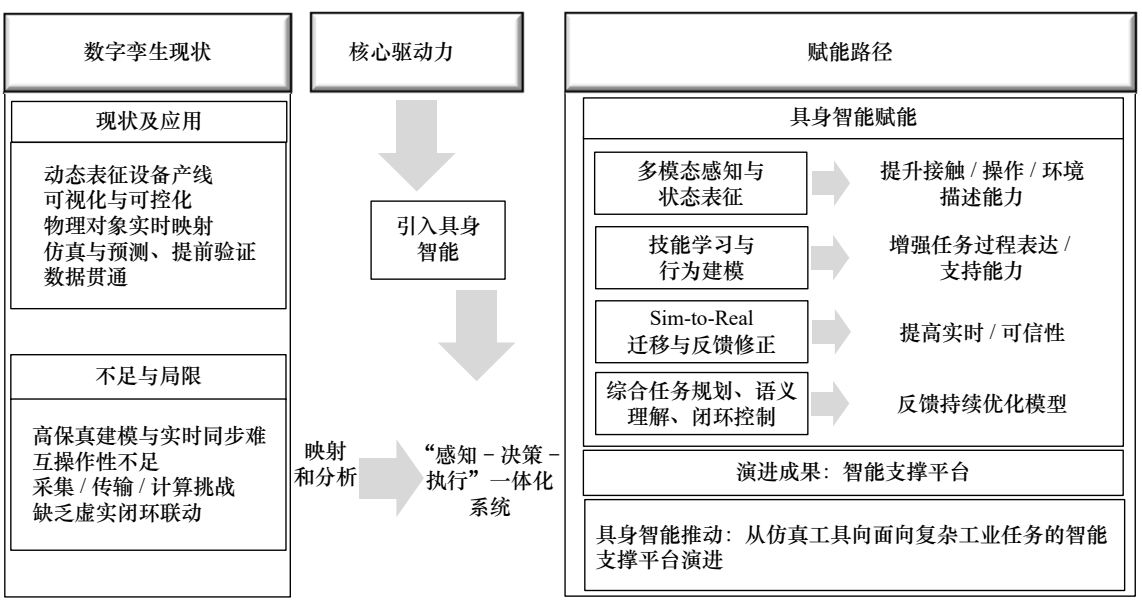


图7　具身智能赋能数字孪生

表1 工业具身智能典型应用场景的任务特性与工程指标对比

典型场景	核心痛点	典型任务	输入模态	所需具身能力与关键技术	控制要求与安全边界	工程评价指标	部署落地难度
人机协作	动态避障与意图识别滞后，共享空间危险系数高	连续轨迹协同、重载移载、工具递送（连续多自由度空间）	视觉、语音、人体骨骼关键点、关节力矩	跨模态语义理解、动态时空避碰规划、“视觉–语言–动作”模型	动态避碰机制、符合 ISO/TS 15066：2016 的安全限速与毫秒级力阈值硬约束	意图识别准确率、协作节拍缩短率、碰撞响应延迟（<50 ms）	强交互下的实时动态安全评估模型缺乏工业级安全认证
柔性生产与精密装配	来料公差波动，微小接触状态难以数学表征	轴孔高精度插接、柔性线缆布线、曲面打磨（微米/亚毫米级约束空间）	高分辨率3D视觉、高频六维力、高精度位姿	高频微小特征识别、阻抗/导纳控制、模型预测控制	亚毫米级轨迹跟踪、高频力位混合伺服、防压溃力矩上限保护	一次装配成功率（>99.9%）、插入力峰值控制精度、作业节拍	复杂非线性接触动力学（摩擦突变、刚度突变）难以高保真建模
物流仓储	非标/软包物料抓取难，多机交叉调度易拥堵死锁	复合大范围移动、动态目标追踪抓取、异形件拆垛（全局二维（2D）移动+局部3D操作）	激光雷达点云、深度视觉、语义地图、里程计	全局多智能体协同调度、仿真到现实的迁移、合成数据泛化抓取	全局动态重规划、局部平滑避障、底盘与机械臂耦合防倾覆控制	动态抓取成功率、多机协同空间利用率、系统平均空载率	长时视界下多智能体协同规划算力要求高，异形抓取泛化性弱
智能运维	隐蔽/高危环境依赖人工，状态监测与异常处理脱节	狭小空间穿梭、仪表识读、阀门旋拧操作（非结构化复杂3D空间）	红外热成像、声学阵列、气体传感、高清视觉	多模态融合诊断、检索增强生成故障分析、高自由度运动规划	未知环境自主防跌落、误操作双重数字校验阻断机制、防爆要求	长尾故障检出率、漏报率、闭环处置成功率、恶劣工况续航能力	长尾异常样本极度缺乏导致大模型易产生幻觉，动作可解释性弱
数字孪生	虚实同步延迟高，工艺交互缺乏闭环验证，仿真到现实的迁移难	策略预训练（虚拟空间）、工艺参数在线滚动寻优（全状态映射物理空间）	工业物联网时序流、PLC运行数据、计算机辅助设计（CAD）机理模型	域随机化、物理信息神经网络、参数在线辨识	通信延时严格低于控制系统扫描周期、虚拟策略下发真机前的前置安全校验	模型物理特征保真度、跨域通信时延、策略真机零样本迁移验证通过率	多物理场耦合计算复杂度极高，异构底层装备数据接口协议壁垒深

生命周期链路，从系统级层面推动工业具身智能的工程落地与自主进化。这四大技术单元高度耦合、互为支撑，共同保障了工业制造系统在面对动态扰动与高度不确定性时的智能性、柔性与可靠性。

（一）多模态感知与状态表征

多模态感知是工业具身智能实现精准环境理解和复杂任务定位的先决条件。在工业环境中，多模态感知的技术难点不在于前端传感器阵列的简单物理堆叠，而在于跨模态数据的时空对齐、异构特征空间的冗余信息挖掘及全局状态表征构建[34-36]。真实的工业场景往往伴随更复杂的特征信息，如噪声强、遮挡多、反光和纹理不足等问题，因此必须综合利用视觉（2D/3D）、力觉、触觉、声学、振动、温度以及位置编码等多源异构信息，构建对工件位姿、接触动力学、设备健康度及全局环境的完备描述[37]。传统工业机器视觉与力控系统通常采用决策级后融合，即视觉系统输出位姿坐标，力传感器输出接触力矩，两者在最高决策层进行逻辑判断。这种方式工程实现简单、解耦性强，但在面对高速运动或微小滑移时，极易因两者采样频率的巨大差异错位。目前的典型方法多采用跨模态注意力机制的特征前级融合，将高维的2D/3D视觉特征与低维的时序力矩、振动特征映射至统一的高维隐空间进行联合表征。近年来，基于深度学习[38-42]、自监督学习[43-46]的多模态融合方法在跨模态特征对齐和弱标注学习方面取得进展，极大提升了模型在未见工况下的泛化表征能力。

当前感知的最大瓶颈在于多模态数据的高频时空同步。视觉传感器（通常30~60 fp/s）与六维力

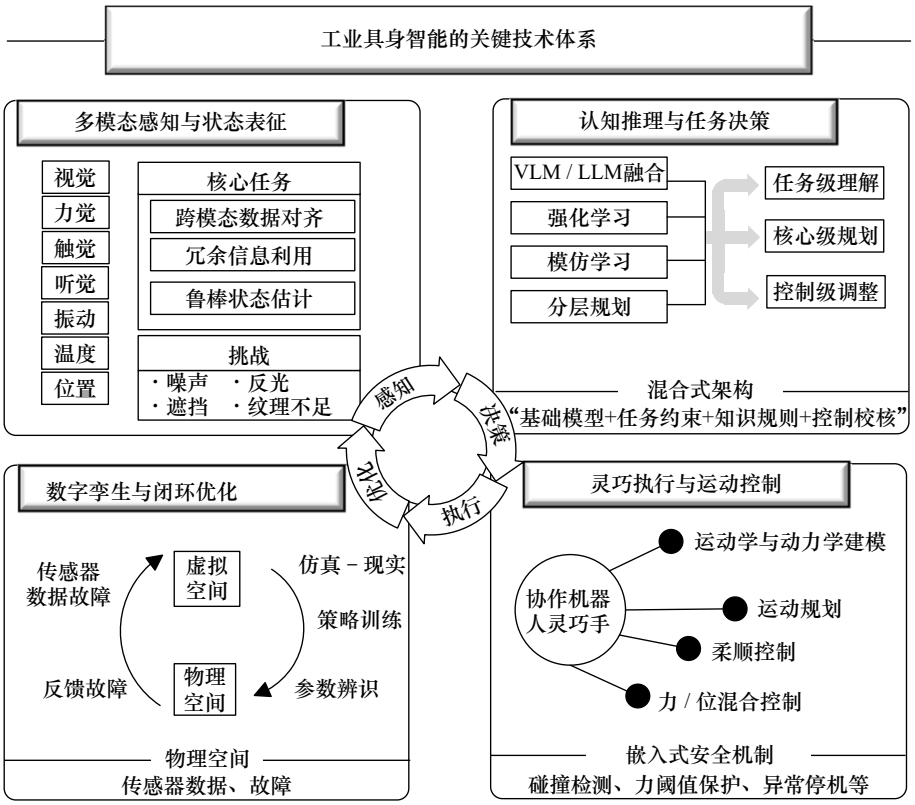


图8　工业具身智能的关键技术体系

传感器（通常1000 Hz以上）存在显著的采样率差异，当发生刚性碰撞或微小滑移时，视觉特征的提取严重滞后于力觉反馈，导致感知表征出现“时间错位”。在真实工程部署中，传感器布设复杂、数据采集代价高、标注困难和实时性约束难，同时工业现场充斥着电磁干扰、油污遮挡与剧烈光照变化等，导致精密光学与触觉传感器极易出现基线漂移和数据缺失，因此，确保感知系统在恶劣工况下连续无故障运行是极大的工程挑战。特别是在强接触操作和高速节拍场景中，感知系统必须与控制环紧密耦合，才能避免“看得见但来不及调整”的问题。

（二）认知推理与任务决策

认知推理与任务决策决定着工业具身智能系统能否将感知结果转化为可执行策略。传统工业自主系统的决策范式是“感知－规划－控制”模块化级联架构，各模块边界清晰，具备极强的可解释性与工程可控性。但该架构的误差会逐级放大，且面对超出预设逻辑的工况（如异常遮挡、非标指令）时系统极易崩溃。具身智能引入的端到端“视觉－语言－动作”模型，打破了这一局限，直接将传感画面与语言指令映射为底层关节力矩或速度。对比模块化架构，视觉－语言模型[47-49]、“视觉－语言－动作”模型[50-53]、大语言模型与强化学习、模仿学习[54-56]、分层规划[57-59]等方法正在加速融合，基础模型为工业具身智能带来了更强的知识迁移和任务泛化能力[60-61]，极大降低了人工构建规则的成本且具备涌现的泛化能力。

然而，认知大模型的引入也带来了性能瓶颈。一是推理延迟，现有百亿参数规模的“视觉－语言－动作”模型的单次推理往往需要数百毫秒甚至秒级，无法满足工业制造毫秒级的节拍要求；二是决策输出的非确定性与幻觉，这种在工业安全中是“零容忍”的。工业决策过程强调安全边界、工艺优先级和异常恢复策略，因此工业场景并不能直接照搬开放数据训练出的通用模型。当前，具身智能的工业落地更适合采用“基础模型+任务约束+知识规则+控制校核”的混合式架构，以兼顾智能性与可靠性。此外，将庞大的认知决策模型部署至产线面临边缘算力极度受限的难题。工业现场通常只能

配备工业工控机（IPC）或轻量级边缘计算盒子，因此如何在保证不损失关键推理精度的情况下，利用模型量化（如INT8/INT4量化）、剪枝与知识蒸馏技术实现本地化轻量部署，且与现有PLC/SCADA系统实现互操作，是系统集成的核心难点。

（三）灵巧执行与运动控制

灵巧执行与运动控制是工业具身智能从“会判断”走向“会完成”的关键。对于抓取、装配、打磨、涂胶、焊接、搬运和在位维护等任务而言，系统不仅需要生成动作，更需要在复杂接触关系和多种扰动下稳定完成动作。针对接触密集型任务（如打磨、精密插接），传统工业控制通常采用PID与阻抗控制或导纳控制，这种方法在面对工业现场复杂的非线性接触动力学（如静摩擦力突变、柔性材料大变形）时，固定参数的阻抗模型往往导致剧烈震荡。面对动态扰动，具身智能执行层引入了基于深度强化学习的策略控制与模型预测控制的结合方法，涵盖预测性碰撞检测、毫秒级力阈值保护、异常状态安全停机、人工接管与历史状态无损回退，综合考虑机构设计、运动学与动力学建模、路径规划、柔顺控制、力位混合控制以及末端执行器适配等问题，保证在高强度重复运行中保持精度、一致性和安全性。例如，装配任务要求机器人可以处理公差、对位误差和接触不确定性；打磨和抛光任务需要根据表面曲率与材料特性动态自适应调整接触力；移动操作任务则需同步处理底盘导航与机械臂末端操作的耦合控制。协作机器人、灵巧手、可变刚度执行机构以及移动操作复合平台的进展，为工业具身智能提供了更丰富的执行载体[62]。

灵巧执行的核心瓶颈在于非线性接触动力学的建模失真。在极小公差插接或精密打磨时，机械臂各关节的静摩擦力突变、连杆柔性形变以及传动背隙呈现强非线性特征，现有的控制算法在应对这类刚度瞬变问题时，极易产生超调或震荡。

（四）数字孪生与闭环优化

数字孪生为工业具身智能提供了一个全息、可计算且具有物理一致性的平行世界，是连接虚拟空间与物理空间的重要桥梁。在构建虚拟映射底座时，传统仿真软件（如基于有限元分析或刚体动力学引擎）需要依赖极其精确的先验参数（如质量、摩擦系数、材质刚度），计算耗时长，只能用于离线工艺规划。而工业具身智能要求的数字孪生必须支撑在线策略迁移。因此，当前的数据孪生前沿方案深度融合了物理信息神经网络与残差物理模型，利用真实产线采集的稀疏传感数据在线反向逼近未知的动力学残差，使虚拟引擎能够自适应地补偿诸如齿轮磨损、温度漂移等带来的物理变异，实现数字孪生体的在线演化。数字孪生不仅是监测工具，更是支撑训练、部署和持续优化的重要基础设施，通过融合CAD/计算机辅助工程机理模型与多传感时序数据，构建包含设备动力学、工艺流场、产线物流与空间干涉的数字映射，可以在虚拟空间中进行状态监测、过程仿真、策略验证和性能评估，再将结果反馈到物理系统，实现虚实协同优化[63-65]。

然而，当前仿真引擎（如Isaac Sim、MuJoCo等）在渲染刚体运动方面已较成熟，但在柔性材质（如线缆布线、织物抓取）的大变形动力学以及多物理场（热、力、电磁）的实时高频耦合渲染方面，计算耗时呈指数级增长，保真度面临极大瓶颈。在真实产线部署闭环优化系统时，不同设备厂商的底层协议相对封闭，数据采集的碎片化与工业通信标准（如OPC UA、MQTT）不统一，使全局高保真孪生底座的构建面临极大的协议转换工作量。此外，将虚实迭代后的新策略在线热更新到正在运行的生产设备中，并确保过渡阶段的绝对平稳，是数据孪生落地的一大挑战。

五、工业具身智能技术面临的应用挑战与未来发展方向

（一）工业具身智能技术面临的应用挑战

1. 仿真–现实迁移鸿沟

仿真训练是降低工业具身智能训练成本的重要手段。真实工业环境中的非线性接触动力学、动态摩擦特性、柔性材料形变（如线缆、织物）、传感器噪声和设备磨损难以在现有物理引擎中被高保真刻画，导致策略空间产生严重偏差。例如，具身智能体在仿真环境中成功率达95%以上的策略，部署到真实产线后，往往面临动作迟缓、接触抖动甚至任务失败。因此，需进一步研发融合数据驱动与物理机理的神经符号渲染技术，利用域随机化增强策

略鲁棒性；同时，在真机部署阶段引入少样本在线强化学习进行残差补偿校准。

2. 高质量多模态工业数据集匮乏

工业具身智能高度依赖多模态数据，但与互联网开源数据不同，工业场景数据带有强烈的企业专有性，多模态数据存在时序同步、空间配准难度大、且缺乏长尾罕见故障样本。数据问题不仅制约模型训练，也影响系统评测和迁移复用。在实际部署中，当多模态数据的单一模态失效时（如视觉系统在低光照条件下），需要考虑如何确保系统的鲁棒性，视觉数据的高维度与力觉数据的低维度如何匹配，不同采样率的传感器数据如何同步等难题。目前，我国在底层算法方面与国外差距逐渐缩小，但开源工业具身数据建设严重滞后。因此，亟需由行业龙头牵头，建立兼顾隐私安全的联邦学习机制，并利用生成式人工智能与工业元宇宙技术合成高质量的边缘工况增强数据，构建标准化的大规模工业具身数据体系[66]。

3. 系统鲁棒性、泛化能力与安全性不足

当前，具身智能系统在实验条件下可展示较好的性能，但工业生产对稳定性和可预测性的要求远高于服务场景，在真实产线的长周期运行中，面对光照突变、非标来料或偶发扰动时，极易出现动作迟缓、接触震荡等性能衰减现象。这是由于基础模型的概率生成机制与工业控制的确切性要求存在底层错位，基础模型引入的幻觉、策略不可解释性和边界条件识别不足，会带来工业安全风险[67-68]，同时，真实设备的长时间运行必然伴随传感器基线漂移、机械传动背隙磨损等非平稳动态变化，这打破了预训练模型所依赖的理想状态假设，导致系统在面对硬件漂移与环境扰动时泛化能力骤降。因此，在控制算法层，需将机械运动学死区、最大接触力矩阈值、绝对禁飞区等物理安全边界作为先验知识进行硬编码，形成大模型指令下发前置的绝对校验门槛；在决策评估层，需引入基于多模态异常检测的实时不确定性评估机制，当模型输出的置信度低于安全阈值时，自动触发系统的“可降级运行”模式（如自适应切换至低速阻抗控制或安全暂停）；在工程部署上，应固化包含毫秒级异常状态急停、无损历史状态回退以及“专家在环”一键接管在内的立体化责任追溯与安全恢复机制，确保系统在未知工况下的绝对可控。

4. 系统集成复杂与工程部署成本高

工业具身智能的工程部署涉及机器人本体、传感器、控制器、边缘计算平台、通信网络和工业软件系统等多层要素，系统集成复杂、调试周期长、初期投资高。对于制造型企业而言，缺乏清晰的投入产出路径和可复制方案，导致技术采纳门槛极高，难以实现规模化推广。此外，与MES、ERP、PLC、SCADA等既有系统的兼容问题，也会显著增加落地难度。因此，具体到工程部署，在技术架构层面，应加速推动工业具身智能软／硬件接口与通信协议（如OPC UA、数据分发服务、时间敏感网络等）的标准化建设，构建物理层解耦、功能层模块化的“即插即用”中间件平台，降低跨系统联调壁垒；在商业模式与部署策略层面，需通过提供标准化的通用硬件载体与“云边协同”的订阅式工业技能包，将重资产的定制集成转化为低门槛的按需配置，从而为中小企业打通普惠式的技术采纳通道。

5. 标准体系与测评框架尚不完善

目前，工业具身智能仍缺少统一的数据规范、接口标准、测试流程和评测基准。不同厂商和研究团队在任务定义、数据格式、模型输入输出和安全验证方法上差异较大，不利于结果比较、系统复用和产业协同。尤其是在工业场景中，单纯依赖学术环境中的基准测试难以真实反映节拍、稳定性、故障恢复和全生命周期成本等核心指标。未来需要构建面向工业任务的多层级评测框架，将算法性能、控制质量、作业安全、部署效率和经济性纳入统一评价体系，并推动接口规范和测试标准的逐步形成。

（二）工业具身智能技术的未来发展方向

面对当前技术挑战和应用需求，工业具身智能正朝着通用化、协同化、自主化方向发展，未来几年有望在以下方面取得重要突破。

1. 面向工业任务的基础模型与通用智能体

未来，发展面向工业任务的基础模型和通用智能体是工业具身智能技术的重要方向。相较于单任务、单场景的专用系统，通用智能体能够在知识层、技能层和策略层实现更大范围的迁移复用，通过构建可覆盖典型工业任务谱系、运用少量数据或少量示教完成适配的模型体系，显著降低模型开发与部署成本。在这一过程中，“视觉–语言–动作”

模型、世界模型和工艺知识库的融合将成为值得关注的方向。通过将任务语义、环境状态、动作表示和约束规则纳入统一框架，有望提升系统在装配、运维和复合操作中的泛化能力。

2. 人机共融的协同智能

工业 5.0 强调以人为中心的智能制造，这意味着未来工业具身智能的发展不应仅以替代人工为目标，而应更加重视与人的长期协同。人机共融协同的核心在于建立更自然的交互机制、更清晰的责任边界和更高效的能力互补模式。未来，系统将更重视对人员意图、操作习惯和现场语境的理解，以实现更顺畅的工位协同和远程协作。在工业实践中，这一趋势将体现在协作装配、远程维护、专家在环决策和技能迁移等方面。通过自然语言、手势、视觉引导和增强现实等方式，人类可以更低门槛参与机器人任务配置与异常处置，机器人则承担高重复、高风险和高负载作业，形成更加稳定的人机协同结构。

3. “云边端”协同与群体智能

单机智能难以满足复杂制造系统的整体优化需求，工业具身智能未来将进一步走向“云边端”协同和群体智能。云端可承担模型训练、知识共享和全局调度，边缘侧负责低时延推理与局部优化，端侧则执行实时控制与感知闭环。通过这一分层架构，系统可以兼顾大模型能力、工业实时性和数据安全要求。在多机器人、多工位和跨产线协同场景中，群体智能将成为提升系统整体效率的重要路径。通过共享知识表征、任务编排和协同策略，不同机器人可在动态订单和复杂环境下完成更灵活的资源分配与流程重构。

4. 虚实融合驱动的自主进化

工业具身智能的长期竞争力将越来越取决于其持续学习和自我优化能力。未来，随着数字孪生、在线学习、增量学习和故障知识积累机制的发展，工业系统将从一次性部署逐步转向持续演化。真实运行产生的数据不再仅用于事后分析，而将作为模型更新、策略修正和工艺优化的重要来源。这一趋势意味着工业具身智能将逐步形成“训练–部署–反馈–再训练”的持续闭环，并在更大范围内实现知识复用和经验迁移。与此同时，自主进化需建立在可控和可验证的前提之上，不能以牺牲安全和稳定为代价。如何在学习能力与工程可控性之间取得平衡，将是未来研究的核心议题之一。

六、推动工业具身智能技术发展的对策建议

（一）加强技术攻关，突破技术瓶颈

具身智能在工业场景中的技术应用已展现出巨大潜力，但我国当前的具身智能发展仍面临瓶颈，如高端六维力 / 触觉传感器高度依赖进口、高质量工业操作多模态数据集（如抓取 / 装配轨迹库）严重匮乏、工业级高保真物理仿真引擎核心技术缺失。未来，需聚焦仿真–现实迁移技术突破，通过研发高精度物理引擎与域自适应算法，有效提高技能迁移效率和成功率；同时，不断加强多模态融合算法研究，重点解决数据异构性与同步性问题，提升复杂环境下的感知鲁棒性。开发标准化、模块化的硬件组件和软件平台，降低系统集成难度与成本；持续强化系统安全性和可靠性设计，通过冗余机制、故障自诊断等技术，确保其在不确定环境中的鲁棒表现。着力构建大规模工业具身数据集，特别是在汽车柔性总装及 3C 电子精密制造等具备海量数据基础的场景优先实现单点突破，以此推动基础模型训练与泛化能力的提升，为工业具身智能的规模化应用奠定坚实基础。

（二）完善产业生态，加快工业化落地

具身智能在智能制造领域的应用和发展，需要构建更为完善的产业生态体系，逐渐推动应用落地。一是促进“产学研用”深度融合，加快科技成果向现实生产力转化，通过构建开放创新的技术平台，鼓励中小企业参与生态建设，并加强知识产权保护与数据安全治理，营造健康可持续发展的市场环境。与此同时，还应注重跨领域人才培养，以满足产业对复合型人才的需求。二是优先选择需求明确、技术成熟度较高的场景（如汽车装配、电子制造、物流仓储）开展试点示范，通过积累成功案例并形成可复制的解决方案，为规模化应用提供实践参考。加快行业标准和评估体系建设，规范技术接口、数据格式和性能指标，促进不同系统间的互联互通，降低技术适配成本；开发适应中小企业需求的轻量化、低成本解决方案，通过简化部署流程与优化成本结构，有效降低技术采纳门槛；加强用户培训和技术支持体系建设，通过系统化培训与全周

期服务，提升企业应用具身智能技术的能力。三是建立第三方检测评估机构，为用户提供客观公正的技术选型参考，从而增强市场信任度与应用信心。

（三）强化政策支持，优化发展环境

科学技术的发展需要国家政策的引导，相关部门需加强顶层设计，制定具身智能产业发展规划，明确技术路线和应用重点，并加大研发投入，支持关键核心技术攻关。通过完善相关法律法规和伦理规范，引导技术负责任发展，鼓励国际合作与交流，融入全球创新网络。工业具身智能作为智能制造的核心技术之一，正推动相关产业的智能化转型和融合发展，形成集创新驱动性、物理依赖性、跨界融合性和系统生态性于一体的新型综合性经济体系。通过持续的技术创新和应用探索，具身智能有望成为推动数字经济高质量发展、促进实体经济与数字经济深度融合的核心力量，为制造业转型升级注入新动能，为制造业高质量转型升级提供重要支撑。

利益冲突声明

本文作者在此声明不存在任何利益冲突或财务冲突。

Received date: April 13, 2026; **Revised date:** June 12, 2026

Corresponding author: Sun Yize is a professor from College of Mechanical Engineering, Donghua University, and a member of Chinese Academy of Engineering. His major research fields include embodied intelligence and intelligent manufacturing. E-mail: sunyz@dhu.edu.cn

Funding project: Chinese Academy of Engineering project “Research on Innovative Development of Functional Textile Materials and Products for Sports Footwear & Apparel in Fujian Province” (2025-DFZD-06)

参考文献

[1] 我国制造业总体规模连续15年保持全球第一[EB/OL].（2025-09-09）[2026-04-08]. https://www.gov.cn/lianbo/bumen/202509/content_7039768.htm.
China’s overall manufacturing scale has remained the world’s largest for 15 consecutive years[EB/OL].（2025-09-09）[2026-04-08]. https://www.gov.cn/lianbo/bumen/202509/content_7039768.htm.

[2] 黄玲玲. 新型工业化进程中数字化制造技术驱动的新质生产力发展研究[J]. 中国经贸，2025（11）：218-219.
Huang L L. Research on the development of new quality productive forces driven by digital manufacturing technology in the process of new industrialization[J]. China Business Update，2025（11）：218-219.

[3] 姚锡凡，马南峰，张存吉，等. 以人为本的智能制造：演进与展望[J]. 机械工程学报，2022，58（18）：2-15.
Yao X F，Ma N F，Zhang C J，et al. Human-centric smart manufacturing：Evolution and outlook[J]. Journal of Mechanical Engineering，2022，58（18）：2-15.

[4] 袁依格，何卓丰，李威，等. 具身智能驱动的智能制造应用发展研究[J]. 中国工程科学，2025，27（3）：67-82.
Yuan Y G，He Z F，Li W，et al. Development of intelligent manufacturing driven by embodied intelligence[J]. Strategic Study of CAE，2025，27（3）：67-82.

[5] 郭宇，龙智新. 智能制造前沿技术概述[J]. 机械制造与自动化，2025，54（3）：1-12.
Guo Y，Long Z X. Overview of frontier technologies for intelligent manufacturing[J]. Machine Building & Automation，2025，54（3）：1-12.

[6] Ren L，Dong J B，Liu S，et al. Embodied intelligence toward future smart manufacturing in the era of AI foundation model[J]. IEEE/ASME Transactions on Mechatronics，2025，30（4）：2632-2642.

[7] 刘馨竹，王亚珅，石晓军，等. 2024年具身智能技术发展分析[J]. 无人系统技术，2025，8（2）：108-122.
Liu X Z，Wang Y S，Shi X J，et al. Development analysis of embodied intelligence technology in 2024[J]. Unmanned Systems Technology，2025，8（2）：108-122.

[8] Fan H L，Liu X，Fuh J Y H，et al. Embodied intelligence in manufacturing：Leveraging large language models for autonomous industrial robotics[J]. Journal of Intelligent Manufacturing，2025，36（2）：1141-1157.

[9] Hu Z H，Yang W K，Zhang M，et al. Toward embodied intelligence：An autonomous exploration system for multimodal UAV[C]// 2024 International Conference on Ubiquitous Communication（Ucom），2024：444-448.

[10] Liu Y，Chen W X，Bai Y J，et al. Aligning cyber space with physical world：A comprehensive survey on embodied AI[J]. IEEE/ASME Transactions on Mechatronics，2025，30（6）：7253-7274.

[11] Wang S L，Zhang J Q，Wang P，et al. A deep learning-enhanced Digital Twin framework for improving safety and reliability in human-robot collaborative manufacturing[J]. Robotics and Computer-Integrated Manufacturing，2024，85：102608.

[12] 兰沣卜，赵文博，朱凯，等. 基于具身智能的移动操作机器人系统发展研究[J]. 中国工程科学，2024，26（1）：139-148.
Lan F B，Zhao W B，Zhu K，et al. Development of mobile manipulator robot system with embodied intelligence[J]. Strategic Study of CAE，2024，26（1）：139-148.

[13] 任磊，董家宝，曾宪超，等. 数字族谱：驱动工业具身智能世界模型[J]. 中国科学：信息科学，2025，55（7）：1748-1765.
Ren L，Dong J B，Zeng X C，et al. Digital genealogy：Empowering industrial embodied intelligence world model[J]. Scientia Sinica Informationis，2025，55（7）：1748-1765.

[14] Hosseinzadeh M，Tanveer J，Rahmani A M，et al. A comprehensive survey of hybrid whale optimization algorithm with long-short term memory：Applications，improvements，and future perspective[J]. Archives of Computational Methods in Engineering，2026，33（3）：3665-3706.

[15] 任磊，王海腾，董家宝，等. 工业智能体：体系架构、关键技术与未来展望[J]. 中国科学：技术科学，2026，56（3）：492-506.

Ren L，Wang H T，Dong J B，et al. Industrial agents：Architecture，key technologies，and future perspectives[J]. Scientia Sinica Technologica，2026，56(3)：492-506.

[16] Jahromi S N，Hajipour E，Ehsan M. Model-based power transformer protection using interactive multiple model algorithm[J]. IEEE Transactions on Power Delivery，2025，40(1)：410-422.

[17] Ma Y S，Che W W，Wu Z G. Event-triggered fully distributed control：A model-free adaptive learning algorithm[J]. IEEE Transactions on Automatic Control，2025，70(8)：5576-5583.

[18] Arraño-Vargas F，Konstantinou G. Modular design and real-time simulators toward power system digital twins implementation[J]. IEEE Transactions on Industrial Informatics，2023，19(1)：52-61.

[19] Aslam A M，Chaudhary R，Bhardwaj A，et al. Digital twins-enabled game theoretical models and techniques for metaverse connected and autonomous vehicles：A survey[J]. Journal of Network and Computer Applications，2025，238：104138.

[20] Black K，Brown N，Driess D，et al. π_0：A vision-language-action flow model for general robot control[PP/OL]. V4. arXiv(2026-01-08)[2026-04-08]. https://doi.org/10.48550/arXiv.2410.24164.

[21] Li H，He X Y，Wu Y L，et al. Digital twin and AI-driven robotic embodied control system：A novel adaptive learning and decision optimization method[J]. Robotics and Computer-Integrated Manufacturing，2026，98：103138.

[22] Li J F，Yang S X. Digital twins to embodied artificial intelligence：Review and perspective[J]. Intelligence & Robotics，2025，5(1)：202-227.

[23] Pretel E，Navarro E，Pujol V C，et al. Digital twins and artificial collective intelligence：Synergies for the future[J]. IEEE Internet Computing，2025，29(1)：75-85.

[24] Sun F C，Liu N J，Wang X Z，et al. Digital-twin-assisted skill learning for 3C assembly tasks[J]. IEEE Transactions on Cybernetics，2024，54(7)：3852-3863.

[25] Yan J H，Liu Z A，Leng J W，et al. Human-centric artificial intelligence towards Industry 5.0：Retrospect and prospect[J]. Journal of Industrial Information Integration，2025，47：100903.

[26] 刘阳，柏永杰，林倞. 面向人机物高效融合与协作的具身智能技术体系[J]. 机器人，2025，47(4)：559–580.
Liu Y，Bai Y J，Lin L. Technological system for embodied intelligence towards efficient integration and coordination of human，robot，and physical world[J]. Robot，2025，47(4)：559-580.

[27] 张弛岳，云静，乌熙靓，等. 自动驾驶中3D目标检测的多模态感知与融合方法综述[J/OL]. 计算机工程与应用，2026-04-03. https://link.cnki.net/urlid/11.2127.tp.20260311.1741.012.
Zhang C Y，Yun J，Wu X L，et al. A review of multimodal perception and fusion methods for 3D object detection in autonomous driving[J/OL]. Computer Engineering and Applications，2026-04-03. https://link.cnki.net/urlid/11.2127.tp.20260311.1741.012.

[28] 刘飞凡，李斯源，刘鹏. 基于多模态大模型的具身智能体任务分解与安全规划[C]//2024中国自动化大会，2024：219-226.
Liu F F，Li S Y，Liu P. Task decomposition and safety planning of embodied intelligent agent based on multimodal large model[C]//2024 Chinese Automation Congress，2024：219-226.

[29] 沈甜雨，陶子锐，王亚东，等. 具身智能研究的关键问题：自主感知、行动与进化[J]. 自动化学报，2025，51(1)：43-71.
Shen T Y，Tao Z R，Wang Y D，et al. Key problems of embodied intelligence research：Autonomous perception，action，and evolution[J]. Acta Automatica Sinica，2025，51(1)：43-71.

[30] 陈文祺，陈佳锋，支鹏翔，等. 视觉－语言－动作模型研究综述：迈向通用机器人[J]. 计算机工程与应用，2026，62(11)：17-40.
Chen W Q，Chen J F，Zhi P X，et al. Survey of vision-language-action models：Towards general-purpose robots[J]. Computer Engineering and Applications，2026，62(11)：17-40.

[31] 张向，黄振宇，谭翠媚，等. 人工智能大模型在工业控制系统中应用现状与发展趋势[J]. 计量科学与技术，2026，70(1)：17-25，35.
Zhang X，Huang Z Y，Tan C M，et al. Application status and development trends of large artificial intelligence models in industrial control systems[J]. Metrology Science and Technology，2026，70(1)：17-25，35.

[32] 任昊，成慧. 具身智能技术发展趋势[J]. 数智技术研究与应用，2025，1(3)：64-76.
Ren H，Cheng H. The development trends of embodied intelligence technology[J]. SmartTech Innovations，2025，1(3)：64-76.

[33] Akbar Gozali A，Faris Ruriawan M，Alamsyah A，et al. Smart train control and monitoring system with predictive maintenance and secure communications features[J]. Transportation Research Interdisciplinary Perspectives，2025，31：101409.

[34] Gusare E，Li H，Huang J B，et al. MAF：Multi-modal adaptive fusion for anomaly detection in attributed graphs[J]. Knowledge-Based Systems，2026，338：115543.

[35] Long K F，Xie G Y，Ma L B，et al. Revisiting multimodal fusion for 3D anomaly detection from an architectural perspective[J]. Proceedings of the AAAI Conference on Artificial Intelligence，2025，39(12)：12273-12281.

[36] Lu S T，Li X Q，Zhang W Q，et al. RoEMF：Rotational embedding multimodal fusion for link prediction[J]. Knowledge and Information Systems，2025，67(5)：4325-4348.

[37] Wang Q，Gao Z J，Mei T Y，et al. Deep learning-based multimodal fusion for improved object recognition accuracy[C]//2024 3rd International Conference on Robotics，Artificial Intelligence and Intelligent Control (RAIIC)，2024：471-474.

[38] LeCun Y，Bengio Y，Hinton G. Deep learning[J]. Nature，2015，521(7553)：436-444.

[39] Mostafa M，Almogren A S，Al-Qurishi M，et al. Modality deep-learning frameworks for fake news detection on social networks：A systematic literature review[J]. ACM Computing Surveys，2025，57(3)：1-50.

[40] Pang G S，Shen C H，Cao L B，et al. Deep learning for anomaly detection：A review[J]. ACM Computing Surveys，2022，54(2)：1-38.

[41] Schmidhuber J. Deep learning in neural networks：An overview[J]. Neural Networks，2015，61：85-117.

[42] Zhang R Z，Li H R，Shen Y X，et al. Deep learning applications in ionospheric modeling：Progress，challenges，and opportunities[J]. Remote Sensing，2025，17(1)：124.

[43] Jing M Y，Zhu Y M，Zang T Z，et al. Contrastive self-supervised learning in recommender systems：A survey[J]. ACM Transactions on Information Systems，2024，42(2)：1-39.

[44] Misra I，van der Maaten L. Self-supervised learning of pretext-

invariant representations[C]//2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2020: 6706-6716.

[45] Yuan M Z, Huang Q, Shen A, et al. Exploring self-supervised learning for 3D point cloud registration[J]. IEEE Robotics and Automation Letters, 2025, 10(1): 25-31.

[46] Zhang Y J, Zhang Y H, Liao W W, et al. Multi-view self-supervised learning on heterogeneous graphs for recommendation[J]. Applied Soft Computing, 2025, 174: 113056.

[47] Hong W Y, Wang W H, Lv Q S, et al. CogAgent: A visual language model for GUI agents[C]//2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2024: 14281-14290.

[48] Yao H T, Zhang R, Lyu H H, et al. Bi-modality individual-aware prompt tuning for visual-language model[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2025, 47(8): 6352-6368.

[49] Zhang Y F, Zhang W B, Zeng Z Y, et al. MapReader: A framework for learning a visual language model for map analysis[J]. International Journal of Geographical Information Science, 2025, 39(9): 1973-2008.

[50] Kim M J, Pertsch K, Karamcheti S, et al. OpenVLA: An open-source vision-language-action model[PP/OL]. ArXiv(2024-09-03)[2026-03-31]. https://arxiv.org/abs/2406.09246.

[51] Li W, Luo D Z, Yang D B, et al. Large language model for Action anticipation[C]// Artificial Neural Networks and Machine Learning - ICANN 2024, 2024: 207-222.

[52] Sautenkov O, Yaqoot Y, Lykov A, et al. UAV-VLA: Vision-language-action system for large scale aerial mission generation[C]// 2025 20th ACM/IEEE International Conference on Human-Robot Interaction (HRI), 2025: 1588-1592.

[53] Tian S, Yu D, Cui L, et al. Semantically supervised SeDINO encoder for visual-language-action model[J]. Applied Sciences, 2026, 16(3): 1464.

[54] Han Y S, Jung J Y. Imitation learning of job dispatching data combining classification and ranking[J]. Computers & Industrial Engineering, 2025, 207: 111305.

[55] Si W Y, Wang N, Harris R, et al. Deep multimodal imitation learning-based framework for robot-assisted medical examination[J]. IEEE Transactions on Industrial Electronics, 2026, 73(1): 928-936.

[56] Zheng B Y, Verma S, Zhou J L, et al. Imitation learning: Progress, taxonomies and challenges[J]. IEEE Transactions on Neural Networks and Learning Systems, 2024, 35(5): 6322-6337.

[57] Ajay A, Han S, Du Y L, et al. Compositional foundation models for hierarchical planning[C]//Advances in Neural Information Processing Systems 36, 2023: 22304-22325.

[58] Wen Z W, Wang Z, Wen X D, et al. A hierarchical planning method for AUV search tasks based on the snake optimization algorithm[J]. Sensors, 2024, 24(22): 7196.

[59] Zhou B Y, Zhang Y C, Chen X Y, et al. FUEL: Fast UAV exploration using incremental frontier structure and hierarchical planning[J]. IEEE Robotics and Automation Letters, 2021, 6(2): 779-786.

[60] Ren L, Wang H T, Dong J B, et al. Industrial foundation model[J]. IEEE Transactions on Cybernetics, 2025, 55(5): 2286-2301.

[61] Brohan A, Brown N, Carbajal J, et al. RT-2: Vision-language-action models transfer web knowledge to robotic control[PP/OL]. arXiv(2023-07-28)[2026-04-02]. https://doi. org/10.48550/arXiv.2307.15818.

[62] Kong L B, Peng X, Chen Y, et al. Multi-sensor measurement and data fusion technology for manufacturing process monitoring: A literature review[J]. International Journal of Extreme Manufacturing, 2020, 2(2): 022001.

[63] 付洋, 曹宏瑞, 郜伟强, 等. 数字孪生驱动的航空发动机涡轮盘剩余寿命预测[J]. 机械工程学报, 2021, 57(22): 106-113.
Fu Y, Cao H R, Gao W Q, et al. Digital twin driven remaining useful life prediction for aero-engine turbine discs[J]. Journal of Mechanical Engineering, 2021, 57(22): 106-113.

[64] 陶飞, 张萌, 程江峰, 等. 数字孪生车间——一种未来车间运行新模式[J]. 计算机集成制造系统, 2017, 23(1): 1-9.
Tao F, Zhang M, Cheng J F, et al. Digital twin workshop: A new paradigm for future workshop[J]. Computer Integrated Manufacturing Systems, 2017, 23(1): 1-9.

[65] 郑杭彬, 刘天元, 郑汉垚, 等. 数字孪生多模态视觉推理的神经-符号系统[J]. 计算机集成制造系统, 2024, 30(5): 1571-1586.
Zheng H B, Liu T Y, Zheng H Y, et al. Neural-symbolic system for multimodal visual reasoning towards digital twin[J]. Computer Integrated Manufacturing Systems, 2024, 30(5): 1571-1586.

[66] Ren L, Dong J B, Zhang L, et al. Industrial metaverse for smart manufacturing: Model, architecture, and applications[J]. IEEE Transactions on Cybernetics, 2024, 54(5): 2683-2695.

[67] 肖峰. 具身智能的三重伦理挑战[J]. 学术前沿, 2025(20): 81-90.
Xiao F. The triple ethical challenges of embodied intelligence[J]. Frontiers, 2025(20): 81-90.

[68] 张樨樨, 贾晓霖. 从人工智能到具身智能: 合法性演化、伦理困境与治理进路[J]. 清华大学学报(哲学社会科学版), 2026, 41(1): 19-34, 241.
Zhang X X, Jia X L. From artificial intelligence to embodied intelligence: The evolution of legitimacy, ethical dilemmas and governance approaches[J]. Journal of Tsinghua University (Philosophy and Social Sciences), 2026, 41(1): 19-34, 241.