中国电子技术标准化研究院
China Electronics Standardization Institute

HUAWEI

2025/4

RESEARCH REPORT ON STANDARDIZATION OF INDUSTRIAL LARGE-SCALE MODELS FOR INTELLIGENT MANUFACTURING

编著 ◎ 中国电子技术标准化研究院
华为技术有限公司等

Research repo

standardization of industrial

large-sca

intelligent manufacturing

面向智能制造的工业大模型标准化研究报告

面向智能制造的工业大模型标准化研究报告

中国电子技术标准化研究院

2025 年 4 月

参编单位（排名不分先后）

中国电子技术标准化研究院
东方电气集团数字科技有限公司
西安航天自动化股份有限公司
深圳市优必选科技股份有限公司
羚羊工业互联网股份有限公司
鼎捷数智股份有限公司
深圳市硅赫半导体有限公司
中国科学院微电子研究所
山东山大华天软件有限公 司
浙江 首席智能技术有限公司
上海文鳐信息科技有限公司
中讯邮电咨询设 计院有限公司
深圳盼月亮创新技术有限公 司
沈阳飞机工业（集团）有限公司
中国科学院沈阳自动化研究所
国家电网有限公司大数据中心
中工互联（北京）科技集团有限公司
深圳市速加科技有限公司
华为技术有限公司
北京思谋智能科技有限公司
中车唐山机车车辆有限公司
科大讯飞股份有限公司
上海逸迅信息科技有限公司
山东省计算中心（国家超级计算济南中心）
凌云光技术股份有限公司
中国电信股份有限公司研究院
中国航天标准化研究所
网智天元科技集团股份有限公司
威艾特科技（深圳）有限公司
卡奥斯工业智能研究院（青岛）有限公司
深圳华大智造科技股份有限公司
星环信息科技（上海）股份有限公司
中国电子工业标准化技术协会
中国石油国际勘探开发有限公司
中国铁道科学研究院集团有限公司

参编人员

范科峰、李瑞琪、韩　丽、卓　兰、蔡宇锋、崔文雅、程雨航、胡静宜、郭小龙、张　祎、王　元、熊建坤、陈柄元、刘　枢、陈鹏光、陶　怡、歹　杰、裴春兴、张秋敏、安在秋、梁乔玲、徐甲甲、刘　影、李腾飞、宋　攀、张　驰、黄伊朴、万　金、金　刚、张　蕊、贾承斌、高兴宇、胡伯源、石海龙、陈如婉、林叠守、洪宝璇、商兴宇、石致远、凌见君、汪云辉、屈亚宁、李建勋、张兴超、张彦兵、陈庆帅、曾繁磊、齐建华、李晋航、韩　鑫、王冬卫、焦继超、周晶川、仲凯韬、芮子文、李　渝、王　超、秦承刚、张学琴、刘本刚、董泽光、李士森、杨自飞、王　挺、洪鹏辉、邓徐韬、凌　乐、王东亮、唐剑飞、夏正勋、范豪钧、邵一凡、祝景阳、梅敬成、赵梦芳、陈振宇、何文渊、李晓雄、马学甲、智　振、李　森、张惟皎、刘志刚、高鹏程、唐永亮、刘小欧、鲁　斌、孙婷婷、孙　鹏、易泰勋、田雪峰、何　强

目 录

目 录

前言

智能制造是制造强国建设的主攻方向，是新时代新征程加快发展新质生产力、推进新型工业化的战略性、引领性任务。工业大模型因其出色的上下文理解、指令遵循、内容生成和场景泛化等能力，已成为推动智能制造的重要使能技术之一。工业大模型与制造装备、工业软件的集成应用，也为人工智能与智能制造的深度融合拓展了空间。

工业大模型在智能制造中的持续发展完善及推广离不开统一、全面和协调的标准体系。然而，随着工业大模型的深入应用，工业大模型标准化工作面临工业场景复杂、数据质量参差不齐、技术更新迭代快、隐私泄露等多方面难题。如何构建普适的、行之有效的标准体系，充分发挥工业大模型的潜力与价值，已然成为当前亟待解决的关键挑战之一。为此，中国电子技术标准化研究院联合各参编单位启动并编制了《面向智能制造的工业大模型标准化研究报告》，具体围绕以下七大主题展开：

1、面向智能制造的工业大模型概述

2、工业大模型现状与挑战

3、工业大模型参考架构

4、工业大模型应用场景

5、工业大模型标准化现状与挑战

6、工业大模型标准体系

7、总结与展望

由于智能制造与工业大模型相关技术发展迅速，研究报告编制时间和作者学识限制，恐有纰漏或不严谨之处，敬请谅解与批评指正。

研究报告编制组

第一章 面向智能制造的工业大模型概述

1.1 智能制造概述

智能制造（Intelligent Manufacturing，IM）是基于先进制造技术与新一代信息技术深度融合的人工智能技术，贯穿于设计、生产、物流、销售、服务等产品全生命周期，具有自感知、自决策、自执行、自适应、自学习等功能特点，旨在提高制造业质量、生产效率效益和提供柔性的先进生产方式。[1]根据智能制造系统架构，智能特征是其重要的描述维度，包括资源要素、互联互通、融合共享、系统集成和新兴业态等5层智能化要求。对于新兴业态的构建，依赖于基于物理空间不同层级资源要素和数字空间集成与融合的数据、模型及系统，实现认知、诊断、预测及决策等功能，且支持虚实迭代优化。此外，智能制造还包括虚实融合、知识驱动、动态优化、安全高效、绿色低碳等特征。

在此背景下，工业大模型具有广泛的应用场景，能够对产品设计、生产制造、质量控制、供应链管理等不同环节进行智能化赋能，因此被视为人工智能应用过程中的关键设施。同时，工业大模型作为人工智能在制造业应用的技术之一，与领域知识的挖掘、融合与应用密切相关，能够支撑认知、诊断、预测及决策等方面的能力建设，促进工业软件、工业装备、生产工艺和服务模式等方面的创新，对于智能制造具有重要意义。

---

[1] 国家智能制造标准体系建设指南（2024 版）（征求意见稿）

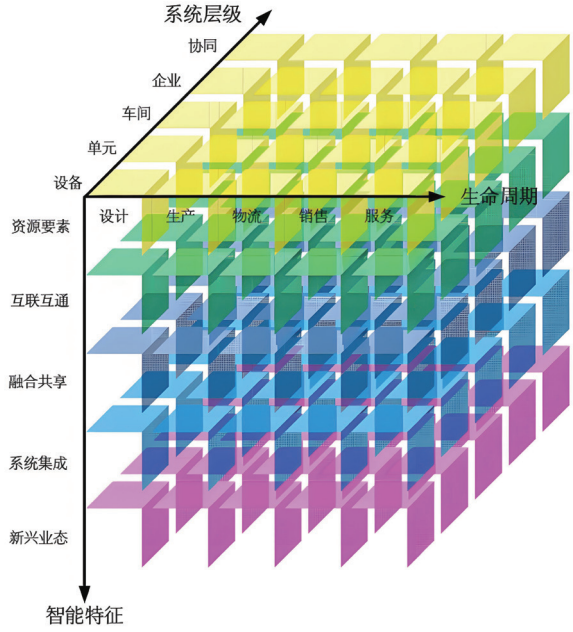

图1-1 智能制造系统架构

1.2 工业大模型的定义和特点

1.2.1 工业大模型定义和内涵

面向智能制造的工业大模型是指在智能制造领域中，利用大规模数据集和复杂的机器学习算法构建的模型。这些模型旨在处理和分析工业生产过程中的大量数据，以实现生产流程的优化、产品质量的提升、资源利用的效率以及维护成本的降低。面向智能制造的工业大模型在通用大模型的涌现能力、通用性和庞大参数规模的基础上，还需要进一步满足生产调度、设备管理、能源管理、安全环保、运行决策等众多制造业专业场景的应用需求，要求其具有较强的专业知识、可靠稳定的输出、严谨的逻辑、安全保密，支持私有化部署并具有较高的性价比，成为可用的“专才”，提供全流程、多要素、多场景的智能化赋能。

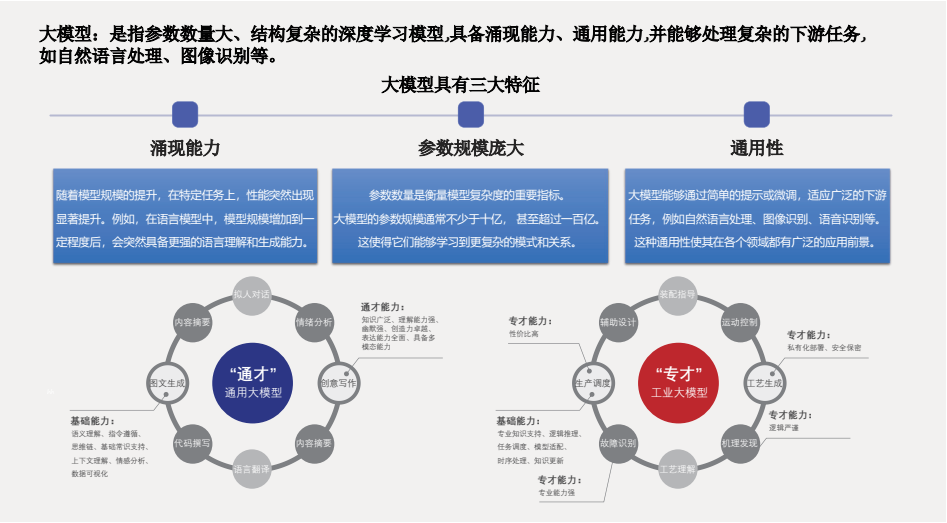

图1-2 工业大模型与通用大模型的区别

1.2.2 工业大模型特点

与通用大模型相比，工业大模型的核心特点包括：

强专业性：既包括工业大模型中所覆盖知识的专业性，也包括了在专业场景中的可用性。

高准确性：经过精心训练和调优，工业大模型能够在特定任务上达到非常高的准确度，满足工业领域对高精度的要求。例如，在设备故障预测、生产流程优化等方面，工业大模型能够提供可靠的结果。

高可靠性：以“零幻觉”为目标，工业大模型在设计和训练过程中特别注意避免幻觉现象，即模型生成的内容与现实不符。这在工业应用中尤为重要，因为错误的预测和决策可能会导致严重的经济损失或安全事故。高可靠性既包括了输出结果的可靠性，也包括了故障情况的可靠性。

可解释性：工业场景中决策指令的下达，需有明确和可辨别的依据，因此对工业大模型的生成结果可解释性提出了较高要求，以便用户能够理解模型的推理和决策过程及其依据。

高稳定性：由于工业大模型在内容生成过程中存在随机性，将会对工业大模型在工业现场的应用带来扰动。稳定性既包括模型自身的稳定性、输出结果的稳定性，也包括模型所提供服务性能的稳定性，例如：输入数据和训练数据的动态更新、操作人员的变化不影响模型输出的准确性等。

高实时性：工业现场设备、生产线及业务软件的运行具有严格的节拍和时间响应要求，要求工业大模型能够快速和及时地完成输出生成。

可集成性：工业大模型需能够与装备、软件、业务系统、企业已有数据库和知识库实现集成，以便支撑制造系统的持续拓展。

安全性：工业大模型在处理敏感数据时，必须确保数据的安全性和隐私保护。此外企业需要建立完善健全的安全政策和控制措施，防止数据泄露和非法访问。

可信赖性：工业大模型的可信赖性主要体现在其高准确性和可解释性。通过对大量行业特定数据的深度学习和分析，工业大模型能够提供可靠的预测和决策支持。此外，其透明的决策过程也有助于增强用户对模型输出的信任，从而确保在复杂工业环境中的有效应用。

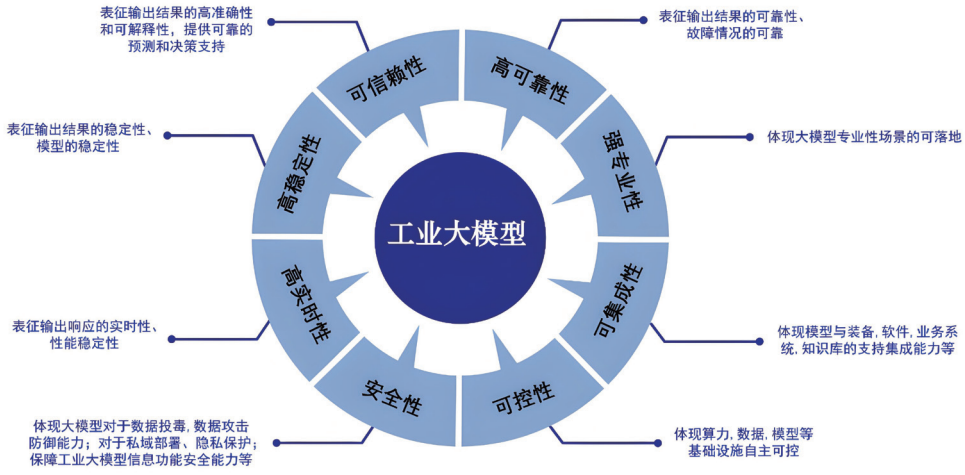

图1-3 工业大模型的核心特点

1.3 智能制造与工业大模型的关系

1.3.1 智能制造推动工业大模型发展

（1）智能制造为工业大模型提供了坚实的数据支撑

智能制造通过物联网、传感器、云计算等设备和设施，获取了海量的实时数据。这些数据不仅涵盖了生产过程中的各个环节，还反映了设备状态、产品质量、企业运营、供应链管理等多个维度的动态信息。而且，制造业企业通过长期的生产实践，积累了大量的图片、图纸、工艺参数等数据资源，均为工业大模型的训练、微调和优化提供了数据支撑。

a. 模型训练与学习：企业内积累的各类型数据可作为大规模数据集提供丰富的训练样本，使得工业大模型能够更准确地对特定场景进行预测和决策。例如，通过分析历史生产数据，模型识别出潜在的故障模式，从而提前预警，降低停机风险。

b. 数据处理能力建设：在智能制造推广和建设过程中，制造业企业不断强化自身的数据治理和安全防护能力，数据存储、清洗和结构化水平逐步提升，并进一步树立了数据质量评价、控制和安全防护意识。这为工业大模型在企业内应用实施过程中所需的标准化、规范化数据处理和接入奠定了基础，降低了前期数据治理难度和成本。

（2）智能制造为工业大模型提供了丰富的工业知识

除了海量的数据，智能制造还促进了制造业企业和行业内工业知识的系统化收集与整理。而且，工业知识图谱作为一种有效的信息组织方式，将分散的知识进行结构化，从而形成一个全面的工业专业知识网络，在部分制造业企业也实现了应用，可用于弥补工业大模型在领域知识和场景知识的不足。而且，工业大模型自身也可以作为数据和知识之间的桥梁，进一步加速工业数据中工业知识的获取速度，为决策提供智能化支持。

（3）智能制造为工业大模型提供了丰富的数字化、网络化软件与装

备环境

依托智能制造改造升级，企业以提高生产效率、降低成本、优化资源配置为目标，已建设了一批数字化车间、智能工厂，并部署了一系列工业软件和装备，例如MES（制造执行系统）、ERP（企业资源规划）、工业机器人、智能采集系统和执行装备等。这些工业软件和装备为工业大模型的后续应用提供了良好的数字化和网络化环境，便于工业大模型与软件和装备的集成，获取工业现场的实时数据。例如：通过实时数据采集与分析，工业大模型能够获取来自生产现场的关键数据，从而实现对生产过程的动态监控和优化。这种实时数据流动使得大模型能够在决策支持、故障预测及资源调度等方面发挥巨大作用。

（4）智能制造为工业大模型提供了大量成体系的优质标准

自2015至2024年，工业和信息化部、国家标准化管理委员会已组织制定了四版《国家智能制造标准体系建设指南》，建立起涵盖基础共性、关键技术、行业应用等标准化方向的智能制造标准体系。截止到2024年，国家标准共计发布400余项；到2026年，预计将制修订100项以上国家标准、行业标准，构建适应新型工业化发展的智能制造标准体系。遵照国家智能制造标准体系建设指南，船舶总装、建材、石化、有色、钢铁、纺织等14个行业在智能制造企业典型经验和应用实践的基础上，进一步开展了细分行业标准体系建设，形成了“国家+行业”的标准体系支撑架构。

上述标准形成了面向场景的标准群，覆盖了数字化车间、智能工厂、信息安全防护、集成优化、装备互联互通、数字化仿真、工艺设计数字化、生产计划优化、质量管控、物流仓储、大规模个性化定制、远程运维等16个具体场景的“标准群”。这都为工业大模型在业务系统和装备集成、数字化车间与智能工厂场景应用、质量管控与安全防护等方面提供了基础指导，降低了探索成本。而且，生产约束、时序要求、接口规范等标准也为工业大模型在不同环境和条件下的一致性和可靠性优化提供了参考。

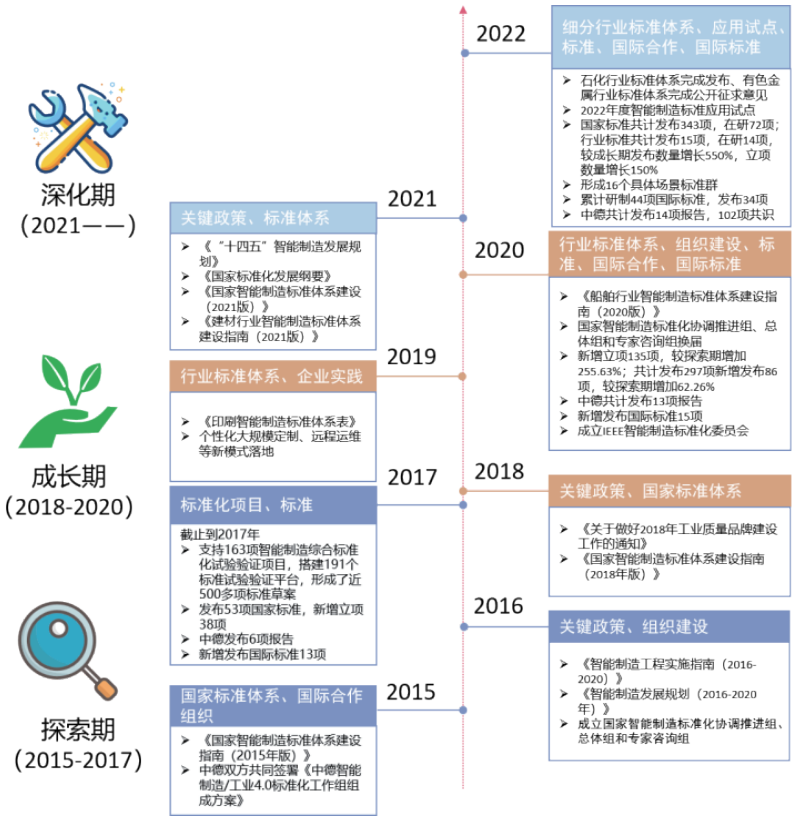

图1-4 标准化主要工作及成效

（5）智能制造为工业大模型奠定了企业智能化转型升级的思想准备和人员储备

《“十四五”智能制造发展规划》中提出：到2025年，规模以上制造业企业大部分实现数字化网络化，重点行业骨干企业初步应用智能化；到2035年，规模以上制造业企业全面普及数字化网络化，重点行业骨干企业基本实现智能化。此外，根据智能制造成熟度自诊断结果，随着智能制造的持续推进，达到智能制造成熟度四级及以上企业（智能化升级）的企业达到了自诊断企业的6%，智能化转型升级已成为制造业企业的重要共识。

智能制造的推进也需要大量的专业人才，包括了智能制造工程师、大数据工程师、数据分析师、物联网和云安全人员、网页开发人员等。而且，部分企业为了推动智能制造的实施，成立了对应的部门或子公司。这些人才储备和机构建设为后续工业大模型的开发和应用提供了人力资源保障。

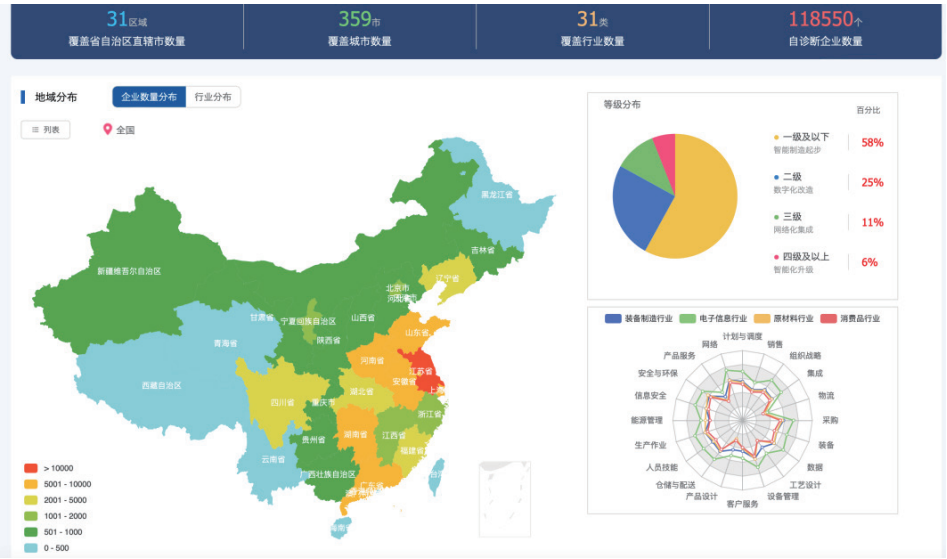

图1-5 我国智能制造成熟度分布情况[2]

1.3.2 工业大模型赋能智能制造

（1）工业大模型为智能制造提供了更强的专业知识迁移能力

工业大模型通过其深度洞察、泛化能力、迁移学习、多模态处理和行业知识整合等能力，为智能制造提供了强大的专业知识迁移能力，这不仅增强了智能制造的适应性和灵活性，也为工业智能化的深入发展提供了坚实基础，更好应对复杂多变的市场需求和生产环境。例如：通过工业大模

[2] 智能制造评估评价公共服务平台 https://www.c3mep.cn/

型上下文语义理解能力，更深层次地洞察工业领域复杂问题，并结合海量数据，从数据中挖掘和总结潜藏的规律和趋势。这使得大模型在研发设计领域能够为产品设计提供精准的创新思路，在经营管理领域提供对生产流程和供应链的管理监控与智能优化。

工业大模型专业知识迁移能力也能够为人才培养提供支持，提升对人员需求变化的适应性。例如，结合大模型生成的培训材料和技术文档，帮助员工快速掌握新的技术和知识；通过虚拟仿真和智能辅导，为员工提供个性化的培训方案，提升其专业技能。

（2）工业大模型为智能制造提供了更灵活的智能化工具

在研发、生产、供应链、运营管理等各环节，工业大模型与传统机器学习模型相比，能够在少样本或零样本情况下为制造业企业提供智能化工具原型开发和业务流程验证，极大减少开发验证周期。例如：在提升研发设计效率方面，工业大模型能够根据用户需求交互，提供产品工程图等可视化文件的辅助生成，提升产线的智能化水平和研发设计的效率。此外，工业大模型依托其对多模态数据的处理和增强能力，能够有效检测出不同生产场景中质量缺陷、人员违规操作、零部件装配误差等风险隐患。

（3）工业大模型为智能制造提供了更强的辅助决策能力

工业大模型结合深度学习、大数据分析、语音识别等技术，可以理解和处理基于语言描述的复杂任务指令，简化语义转换和识别流程。同时，工业大模型也能够模拟领域专家的决策思维，从多个维度分析问题，呈现思考推理过程，提供更加智能化和精准化的类人决策支持，以帮助探寻最优解决方案。例如，在化工行业，大模型可以通过分析已知材料的分子数据，结合专家思考模式，找到适合目标场景的最优候选材料，并生成催化剂分子设计方案。

（4）工业大模型为智能制造提供了更泛化的场景支撑能力

工业大模型的出现一定程度上解决了传统模型或服务不易实现跨模

态、跨领域应用的问题。企业平台的MaaS层可在多种场景下直接为用户终端提供高质量的大模型服务，通过调用API，结合特定业务场景的解决方案对数据进行加工应用，推动企业个性化应用业务的部署、优化和升级。而且，工业大模型通过迁移学习等方式，可在不进行重新训练的条件下适应多种工业场景，降低对工厂样本数据量和本地化训练的依赖度。

（5）工业大模型为智能制造提供了更丰富的多模态数据和图文处理及生成能力

工业企业除了生产环节时序数据外，在设计、服务、运营管理等环节涉及大量的语音转换、图像识别、文档处理、设计文件编制等工作，因此对工业大模型的语言理解能力和图像生成能力需求较高。而且，工业大模型的创意生成、三维可视化呈现和代码辅助编写也可帮助研发设计人员打破思维瓶颈，激发方案构想，降低工作强度。此外还可通过获取与分析生产过程中产品的实时照片或声音，识别产品中的微小缺陷，并进行修补和完善。

（6）工业大模型为智能制造提供了更强的装备感知、规划和执行能力

工业现场装备和生产线通过多类型的传感器组合，可以采集更为全面的现场信息，涵盖视觉、力觉、声音、时序数据等，进而能够获得比单一传感器更准确的信息。工业大模型可对多类型、多模态传感器数据进行融合，获得更强的综合感知能力，并与工业机器人等装备的规划、控制和执行部件集成，构建装备大脑，以实现更精确的动作控制。

第二章 面向智能制造的工业大模型现状分析

2.1 政策现状

2.1.1 国际政策现状

在全球化竞争日益激烈的背景下，德国于2013年在汉诺威工业博览会上首次提出了“工业4.0”这一革命性概念。美国先后通过“先进制造伙伴计划”（AMP）和“国家制造创新网络”（NNMI）等项目，推动了制造业的创新和转型。这些项目旨在通过公私合作，加强研发和人才培养，推动新技术的商业化。日本通过“社会5.0”战略，推动制造业与信息技术的融合，以实现更加智能和可持续的制造系统。日本政府支持企业采用物联网、大数据和人工智能等新技术，提高生产效率和产品质量。欧盟通过“地平线2020”计划，支持智能制造和工业大模型的研发，同时欧盟还推动了“工业5.0”概念，强调人与机器的协作，以及对环境和社会的责任感。根据IDC的报告，电力、采矿、油气、半导体、汽车、消费品等行业头部企业对工业大模型的探索进展较为深入，报告指出，工业大模型的应用已经渗透到工业的多个业务环节，并在众多业务流程的功能点上形成一些应用，相比传统AI场景显得更加碎片化。

美国对人工智能的政策及发展规划起步较早，其中2016年发布的《为未来人工智能做好准备》为其奠定了发展基调。2023年，美国白宫更新了《国家人工智能研发战略计划》，明确强调投资下一代人工智能技术以推动负责任的创新，并提出了九项主要战略来确保美国在人工智能领域的领先地位。2024年，围绕人工智能技术出口管制、风险框架建设及标准化教育等方面，美国政府进一步强化国际竞争力，通过一系列法律和政策措施，推动关键技术标准的预研与全球推广。在标准化领域，美国以NIST为核心，提出并制定了广泛的人工智能标准、建立技术基准、参与国际标

准化制定等多项任务，同时通过《关键和新兴技术国家标准战略实施路线图》等一系列文件，推动了标准化影响力的提升。

欧盟在人工智能领域强调“以人为本”的发展理念，自2018年《人工智能时代：确立以人为本的欧洲战略》发布以来，政策侧重于营造公平、透明和安全的发展环境。2024年，欧盟通过《人工智能法案》，以四个风险等级监管AI系统，并推出了《生成式人工智能治理框架》和《AI数据治理条例》，以强化数据使用规范与隐私保护。此外，欧盟积极发布《人工智能伦理指南》，明确AI开发和使用中的透明性与公平性要求，为全球AI治理提供了创新思路。这些政策与法规体现了欧盟对人工智能的系统性监管和推动创新并存的战略。

日本的人工智能政策近年来逐步聚焦于可信赖AI的发展与国际规则的主导权。自2022年《人工智能战略2022》发布以来，日本通过年度综合创新战略及白皮书，不断明确其AI技术发展的方向和路径。2024年，日本进一步强调国际竞争力，提出以扩大尖端技术投资为核心，设立“AI制度研究会”并开展立法研究。在标准研制方面，日本主导了“广岛人工智能进程”，提出全生命周期监管与数字水印技术等11项指导原则，同时发布了《人工智能运营商指南（草案）》和《人工智能红队测试方法指南》，以应对生成式AI的技术变化，并确保AI系统的安全性与问责机制。

2.1.2 国内政策现状

2020年，中国国家发改委发布了《关于加快推进数字经济发展的实施意见》，强调加快人工智能、大数据等新一代信息技术发展的重要性，并明确提出推动产业数字化转型的目标。随着“十四五”规划的启动，中国进一步深化智能制造发展战略，在2021年发布的《“十四五”智能制造发展规划》明确提出：到2025年，规模以上制造业企业将实现数字化网络化，重点行业骨干企业初步实现智能化，为2035年全面智能化奠定坚实基

础。这一规划的实施将极大提升中国制造业的创新能力和国际竞争力，推动中国从“制造大国”向“制造强国”跨越。

与此同时，同年发布的《“十四五”工业绿色发展规划》聚焦工业的绿色低碳转型。规划以绿色发展为主线，推动工业结构优化升级，加快工业节能降碳，提高资源利用效率，推动清洁生产改造，提升绿色低碳技术水平。规划的实施将促进工业发展与生态环境保护的协调统一，构建绿色、低碳、循环的工业体系，实现工业的可持续发展。

《北京市加快建设具有全球影响力的人工智能创新策源地实施方案（2023-2025年）》以及《北京市促进通用人工智能创新发展的若干措施》明确了推动人工智能产业大模型应用的发展方向和具体目标。这些措施旨在提升算力资源统筹供给能力、高质量数据要素供给能力，系统构建大模型等通用人工智能技术体系，并推动通用人工智能技术创新场景应用。

《上海市推动人工智能大模型创新发展若干措施（2023-2025年）》旨在深入贯彻国家发展新一代人工智能的战略部署，提出了支持大模型创新能力、提升创新要素供给能级、推进大模型创新应用、营造一流创新环境等措施。该政策将推动上海大模型创新发展，营造通用人工智能创新生态，加快打造世界级人工智能产业集群。

深圳罗湖区政府推出了49个应用场景，旨在全面提升该区域的智能化水平，将AI技术的应用延伸至各个行业。这些场景涵盖了社区治理、教育、医疗、文旅和金融等多个细分领域，推动了人工智能技术的广泛应用。

此外，中国国家自然科学基金委员会、中国工业和信息化部等部门在2021年联合发布了《人工智能领域科技计划指南》和《关于促进人工智能和实体经济深度融合发展的指导意见》，进一步强调对人工智能基础理论、关键技术、应用示范等研究的支持，以及推动人工智能与实体经济深度融合，加快产业智能化升级的愿景。2022年，中国科技部等六部门联合

印发了《关于加快场景创新以人工智能高水平应用促进经济高质量发展的指导意见》，鼓励包括制造业在内的多个重点产业深入挖掘人工智能技术应用场景，促进智能经济高端高效发展。

2023年12月，中央经济工作会议召开，会议再次强调要发挥科技创新引领作用，加快现代化产业体系建设，积极主动适应和引领新一轮科技革命和产业变革。会议还指出，要大力发展数字经济，加快发展人工智能等新兴产业，鼓励绿色低碳产业发展，运用数智技术、绿色技术等先进适用技术为传统产业注入新动能，加快实现转型升级。

2024年7月，北京市发展和改革委员会联合多部门印发《北京市推动“人工智能+”行动计划（2024-2025年）》的通知，该通知强调利用工业企业场景和科研院所的聚集优势，搭建基于大模型智能平台的工业软件与生产制造体系，深度挖掘工业数据潜力，重塑研发生产流程，推动新型工业化发展。

2024年12月，上海市人民政府办公厅印发《关于人工智能“模塑申城”的实施方案》的通知，该通知强调加快构建中文工业通识知识库，推动L1模型研发和超级场景规模化应用，建设模型即服务平台，支持L2大模型池，聚焦产品设计、智能排产、智慧物流等重点场景，培育专业服务商队伍。

综上所述，为响应中央政府的号召，各地政府纷纷采取行动，通过制定一系列政策激励人工智能大模型在产业领域的广泛应用。

2.2 行业现状

2.2.1 工业大模型的行业应用进展概述

自2022年ChatGPT发布以来，全球市场对大模型的需求迅速增长。根据Precedence Research的预测，全球大模型市场规模到2032年将激增至

1181亿美元。据艾瑞咨询预测，2030年中国人工智能生成（AIGC）产业规模也将达到11441亿元。

国内外科技巨头纷纷推出自己的通用大模型产品，例如GPT-4、Gemini、LLaMA、文心一言、通义千问等，促进了大模型在多样化场景中的广泛应用。在智能制造领域，通过结合上述通用大模型和特定工业场景的专有数据，形成了垂直化、场景化、专业化的工业大模型。相比通用大模型，工业大模型具有更高的专业度和落地性，能够实现工业各环节中多场景的精确模拟。在工业大模型的辅助下，企业能够预测生产过程中出现的不确定性问题，提前采取预防措施，减少生产终端和质量缺陷的风险。同时，工业大模型还能够优化资源配置，提高生产效率和产品质量，使企业能够更好地适应市场变化，快速响应消费者需求。

人工智能在工业应用领域，以工业大模型为代表的创新技术不断涌现，实现对传统工业领域的赋能。工业大模型应用方向包括产品智能设计、系统智能人机交互、生产线自我优化、设备预测性维护、质量控制自动化、智能物流规划和智能供应链管理、能源消耗优化等，部分企业已崭露头角，如华为、百度、Salesforce、Authentise、中工互联、羚羊工业、奇智创新等等，竞争格局初步形成，涵盖了研发、设备、生产、管理等不同工业环节，涉及制造、矿山、能源、航天等多个领域。这些大模型主要以大语言模型为主，同时包含了一些专用结构化数据大模型、多模态大模型和机器视觉大模型。未来，在工业领域深耕细作，主打“专而精”的工业大模型将成为新型工业化进程的核心驱动力。

尽管大模型在工业领域的应用已经取得了一些成果，但整体普及率仍然较低。凯捷的统计数据显示，即便是在欧洲、日本和美国等地区的顶级制造企业中，人工智能应用的普及率也仅超过30%，而中国制造企业的人工智能普及率更是不足11%。这一数据反映出未来国家在推动人工智能应用普及方面的巨大潜力和空间。

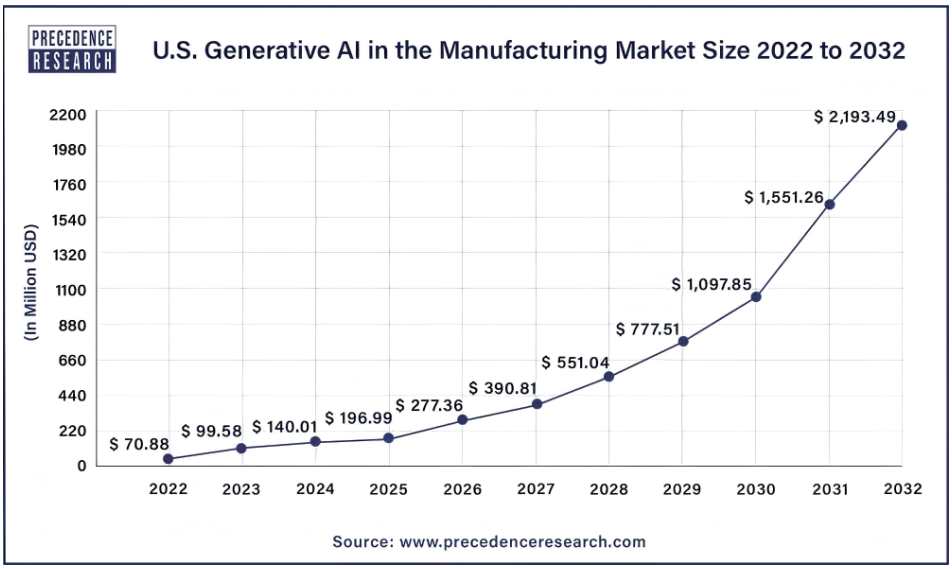

图2-1 美国工业大模型市场规模[3]

2.2.2 工业大模型在汽车行业应用现状

中国汽车工程学会发布《汽车智能制造团体标准体系建设指南》，该指南规划了汽车行业模型选择与训练、模型实施与优化、系统集成、规模与边界、兼容性、交互模式、模型管理、训练数据等标准内容所赋能的汽车行业应用场景。中国人工智能产业发展联盟发布《面向行业的大规模预训练模型技术和应用评估方法 第4部分：汽车大模型》标准，共计3个能力域、10余个能力子域、20余个能力项，围绕汽车生产、研发、使用、销售、售后全环节，形成面向汽车行业的大模型应用成熟度评价方法。

华为公司发布的华为云盘古汽车大模型结合汽车行业的特点和需求，采用汽车行业数据进行训练，进行深度定制和优化，支持汽车行业“研、产、供、销、服”环节，可应用于自动驾驶研发数据管理、企业生产排

[3] 资料来源：precedenceresearch.com

产、产线安全和车辆质检场景中，具备设计文档问答、多模态内容理解与生成、仿真场景生成、代码生成/补全、智能化软件交互集成等功能，加速车企迈向智能化升级。

图2-2 汽车行业大模型特征图

2.2.3 工业大模型在轨道交通行业应用现状

轨道交通行业作为国民经济的重要基础设施，其运维场景复杂，涉及车辆、信号、供电、通信等多维度系统，长期以来面临知识传承困难、运维效率低下等问题。尤其是在故障处理、数据记录与管理等环节，传统依赖经验传递和人工操作的模式已经难以满足轨交行业日益增长的智能化需求。为了解决这些痛点，业内企业积极探索工业大模型技术在轨道交通领域的应用。

上海逸迅发布的轨道交通行业大模型，涉及车辆运维、信号运维、机电运维、公务运维、供电运维、通信运维、客流运维七大板块，这七大板块包含了所有轨交领域的分支方向。目前该模型项目已经完成了部署和实施。该模型拥有庞大的轨交知识库，可以对地铁、高铁、船舶等所有轨交领域以上七个板块的专业问题进行解答，方便新进人员快速上手工作，简

化了老师傅手把手带新人的繁杂过程；除了对专业问题进行解答外，还可以对历史信息进行记录与保存，便于工作人员对历史故障信息、故障处理方法、处理人员等历史记录进行查询，为故障处理提供历史依据，实现轨交系统的智能运维，具有较高的实用价值。

中车发布的斫轮大模型体系涵盖算力、MaaS平台、基模、应用、服务及生态的全链条，其中应用覆盖了设计、制造、运营、维护、安全等制造业各环节。斫轮大模型以“基础大模型、行业大模型、业务大模型、场景大模型”四级模型为支撑，围绕业务全流程、管理全覆盖、客户全周期、产业全领域、行业全生态，推动实现经营效率更高、客户价值更高、治理能力更高、安全水平更高、发展质量更高的智能行业生态，为培育新质生产力，实现装备制造业高质量发展保驾护航。

2.2.4 工业大模型在电力行业应用现状

工业大模型在电力行业的应用里正逐步成为提升效率和实现可持续发展的关键工具。首先，在电力需求预测方面，工业大模型利用历史数据和实时监测信息，能够准确预测用电负荷变化，帮助电力公司更好地调配资源，优化发电和供电计划。这种精确的需求预测不仅提高了运营效率，还能有效降低能源浪费。其次，在智能电网管理中，工业大模型分析海量数据，实时监控电网运行状态，及时识别潜在的故障和风险。这种实时监测与预警机制提升了电网的安全性和可靠性。此外，工业大模型还可以与无人机、摄像头等设备联动，结合视觉大模型，在电力行业安全生产、风电设备巡检等领域发挥关键功效。

最后，随着电力行业向低碳化转型，工业大模型在碳排放监测与管理方面也展现出重要价值。通过分析不同发电方式的碳排放数据，企业能够制定更为科学的减排策略，助力实现环境可持续发展的目标。综上所述，大模型在电力行业的应用，推动了智能化、数字化的发展，为行业带来了更高的效率和更绿色的环境治理。

2.2.5 工业大模型在能源行业应用现状

能源行业是最早尝试开展大模型应用的行业之一，主要应用于能源管理优化、预测需求和提高效率等方面。通过分析大量数据，这些模型能够精确预测能源消费趋势、识别可再生能源潜力并优化供应链管理。同时，它们在故障检测和维护预测方面也表现出色，帮助企业降低成本并提升安全性。整体而言，大模型有助于能源行业实现可持续发展和智能化转型。

科大讯飞羚羊工业互联网平台2024年6月27日发布了能源大模型，该模型以讯飞星火通用大模型为底座，结合能源行业场景需求，打造并实现设备运检、电力问数、电力营销客服、辅助电力交易、新能源功率预测、安全生产等6大应用场景能力，全面赋能风、光、水、火、核、储等领域，为推动能源焕新提供丰富的解决方案。

图2-3 羚羊能源大模型

2.2.6 工业大模型在生物医药行业应用现状

近年来，工业大模型在生物医药领域的应用取得了显著进展。深圳华大智造科技股份有限公司发布的αLab Studio智能实验室助手工业大模型，充分融合了大数据、人工智能与实验室自动化技术，已广泛应用于基

因组学、蛋白质组学、药物研发等关键场景。通过自然语言处理和大语言模型驱动，αLab Studio能够实现智能化实验设计、实验流程自动化以及对多维度数据的深度分析，极大地提升科研效率和实验精准度。

同时，该大模型还支持实验室信息管理系统（LIMS）等基础设施的无缝集成，使实验操作更加高效、标准化和智能化。为了推动工业大模型在生物医药领域的规范化应用，欧洲分子生物学实验室等行业学术组织陆续发布了关于大模型技术在药物开发、基因组研究中的指南和建议文件。这些指南文件的发布，标志着工业大模型在生物医药领域的应用正逐步从实验室阶段向实际产业落地转化，助力精准医疗和科学研究的快速发展。

图2-4 智能实验室助手αLab Studio

2.2.7 工业大模型在高端装备制造行业应用现状

大模型在中国高端装备制造行业的应用逐渐深化，涵盖了设计优化、生产流程管理和质量控制等多个方面。在设计阶段，工程师利用大模型构建知识库，进行复杂数据分析，以辅助产品设计、优化产品结构和性能。这种方法能够显著缩短研发周期，提高设计的精准性，以满足日益增长的市场需求。在生产过程中，大模型通过实时数据监测和分析，优化生产调

度和资源配置，从而提升整体效率，降低生产成本。这种智能化的生产方式，使企业能够更灵活地应对市场变化。

此外，大模型在故障预测和维护方面的应用也日益显著。通过对设备运行数据的持续分析，大模型能够提前识别潜在的故障风险，从而实现精准的预维护，减少设备停机时间和维护成本。这种主动管理策略提升了生产的安全性和可靠性。随着智能制造和工业互联网的发展，大模型在高端装备制造中的重要性将愈发凸显，助力企业实现数字化转型，提升国际竞争力。整体来看，大模型的应用不仅推动了技术创新，还促进了中国高端装备制造行业的可持续发展。

2.2.8 工业大模型在钢铁行业应用现状

2024年中国人工智能产业发展联盟发布《面向行业的大规模预训练模型技术和应用评估方法 第8部分：工业》标准，规定了大规模预训练模型在工业应用的指标要求，主要包括技术支持度和应用成熟度两个维度。同时，该文件旨在指导工业大模型技术方开展产品研发，为工业大模型应用方产品选型、系统建设、应用管理等方面提供参考。

当前的钢铁行业中存在着自动化程度低、控制效率低等问题，这些问题严重影响着钢铁行业的数智化转型。针对这些问题，中国电信结合自身云网融合优势和大模型技术积累，提出了工业质检和节能控制大模型，赋能钢铁行业的各个环节：

由于制造产品过程中产品表面的缺陷不可避免，且传统检测方法样本量少、泛化性差，中国电信提出了工业质检大模型，如图2-5所示。利用大模型的泛化能力，缩短模型适配和优化时间，企业能够实现零样本或者少样本情况的质量检测。

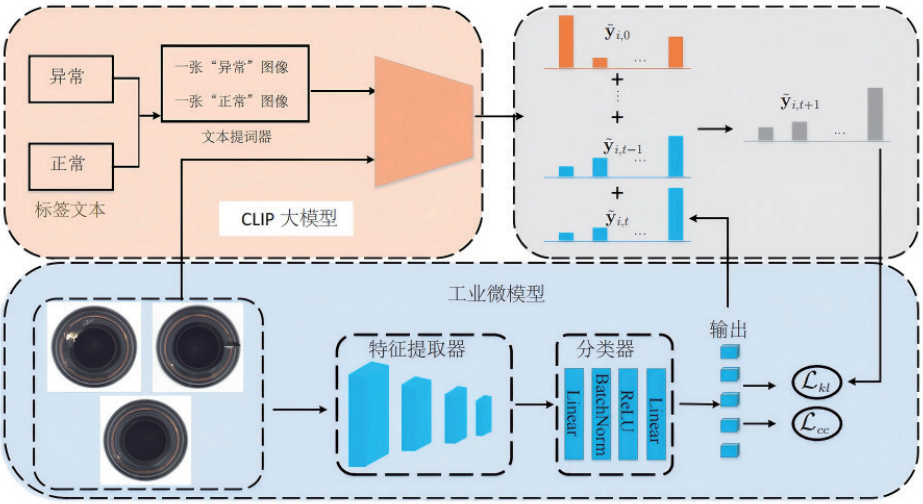

图2-5 工业质检大模型构建方案

钢铁行业中的关键环节（例如：炼铁、炼钢、烧结等工序）需要使用高功率风机进行操作（例如：除尘、供气、冷却等），然而由于非智能化的原因，这些环节往往容易造成大量能源浪费。针对此问题，中国电信面向钢铁行业提出了节能控制大模型，通过引入时序预测大模型协同融合赋能工业应用，可将实施周期压缩10%左右；同时结合时序大模型的预测能力和翼云控的实时响应能力，解决控制响应滞后的问题，将节能率提升3%-5%。

2.2.9 工业大模型在工业企业数字化转型的应用现状

当前的传统工业企业普遍面临数字化水平较低、企业数字化转型人才不足、转型需求碎片化等困境。具体来说，企业在转型过程中需要克服解决方案与需求不匹配、缺乏顶层设计和规划、数据孤岛和数据管理不足等问题。这些问题不仅影响企业的数字化转型，还可能影响企业的长期发展和市场竞争力。因此，企业需要在等多方支持下，制定合理的转型策略，

加强技术创新和人才培养，优化数据管理和安全防护，以实现数字化转型的成功。

在此背景下，中国电信提出了星辰行业大模型。该大模型依托云网融合优势，以工业咨询诊断为基础，向企业提供数字化转型百科问答服务，培育企业转型意识，帮助企业梳理研、产、供、销、服各个环节存在的问题，并给出数字化转型路径规划建议，最后生成诊断报告，助力企业落地改造，推动制造业高质量发展，加快推进新型工业化进程，具体技术流程如图2-6所示。在具体应用场景中，通过对话式问答，结合工业知识库，解答企业生产、经营等过程中存在的问题并给出相应建议，为企业提供推荐产品、解决方案等资源；通过线上的模式对工业企业进行专业的数字化转型咨询，规划转型的路径，目前已支持两化融合评定模型、智能制造能力成熟度模型，在2000+专业诊断报告训练基础上，根据不同诊断模型，输出不同的诊断报告，并根据企业诊断情况，推荐转型资源。

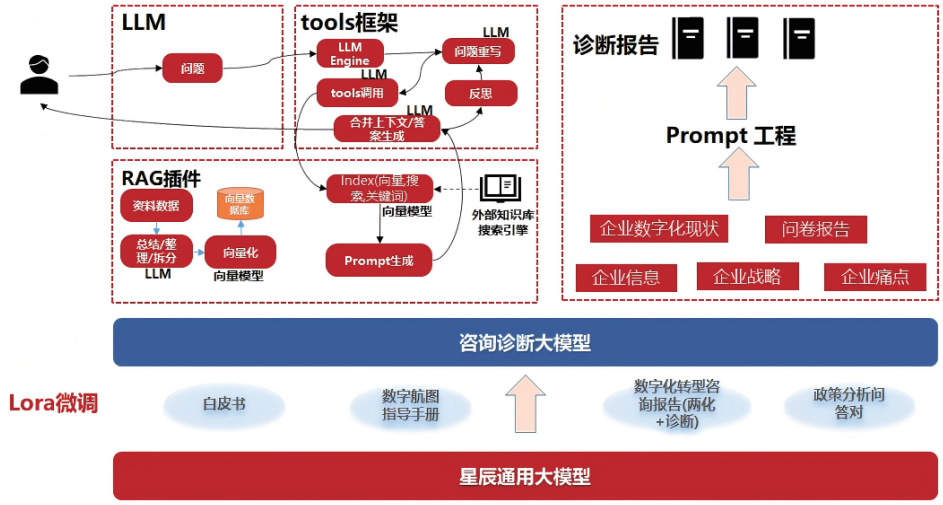

图2-6 工业咨询诊断大模型技术流程

2.3 技术应用现状与挑战分析

当前，工业大模型在智能制造领域呈现多样化的应用形式，除直接应用外，还包括“大模型+工业知识”、“大模型+小模型”、“大模型+装备/应用”、“大模型+工具链”等与外部工具深度融合的模式。不同的应用形式各具特点，适配不同场景需求。本研究报告将围绕现有技术应用现状进行分析，并提出面向智能制造的工业大模型参考架构，为行业标准化发展提供有力借鉴。

2.3.1 大模型直接应用

2.3.1.1 概述

大模型直接应用是指将工业大模型直接用于工业场景，以实现智能分析、优化决策和任务执行。其优势在于通用性与高效性，可快速适配多样化应用场景，显著提升智能化水平。然而，这种应用对模型精度、数据适配性及计算资源要求较高，实际部署中通常结合轻量化和私有化策略，以优化性能并保障数据安全。

2.3.1.2 技术现状

（1）模型训练与微调技术

工业大模型的广泛应用离不开强大的模型训练与微调技术。模型训练通常依赖于大规模数据集和高性能计算资源，而微调技术则使大模型能够针对特定工业场景实现高效优化。在实际应用中，微调技术通过利用预训练模型，借助少量数据进行针对性优化，以适配多样化工业任务。例如，在制造业中，微调技术可快速应用于新产品的缺陷检测任务，不仅显著缩短训练时间，还能提升模型性能。

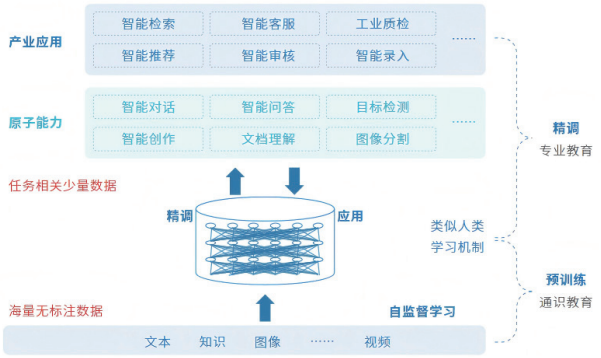

图2-7 预训练与精调联合应用模式[4]

（2）轻量化部署技术

为了满足工业场景中对实时性和低资源消耗的要求，轻量化部署技术成为了大模型应用的关键。通过模型压缩、剪枝、量化等技术，开发者可以将原本复杂的大模型缩减为更轻量的版本，以便在边缘设备或资源受限的硬件上部署。这种技术使得大模型可以在工厂生产线、机器人设备等资源有限的环境中高效运行，实现快速推理和决策。

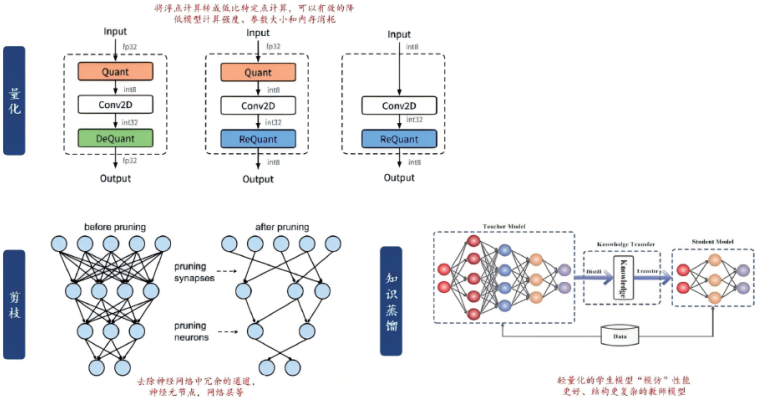

图2-8 模型轻量化的主要三种方式[5]

[4] 资料来源：IDC、百度

[5] 资料来源：华为云官网、华泰研究

（3）私有域部署技术

为保护企业敏感数据并保障工业机密，私有域部署技术正逐步成为工业大模型应用的重要发展方向。通过将大模型部署于企业内部的私有云或本地数据中心，企业能够有效降低数据泄露至公共云平台的风险，同时实现更高水平的数据安全与隐私保护。私有域部署还使企业能够结合自身特定的安全需求和硬件条件，对模型进行深度优化，从而在确保安全合规的前提下充分发挥大模型的效能。例如，许多制造企业选择在本地服务器中部署大模型，以兼顾生产效率提升与数据安全保障的双重目标。

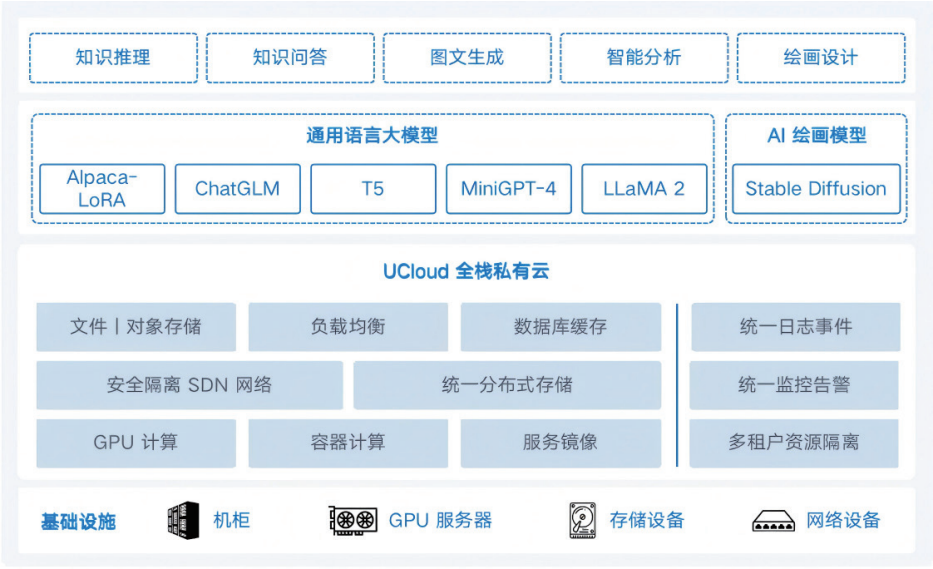

图2-9 UCloud私有化部署方案[6]

2.3.1.3 技术挑战

（1）多模态数据融合难度高

在智能制造中，大模型常常需要结合多模态数据（如二维图像、三维

[6] 资料来源：UCloud 官网

扫描、传感器数据等）进行分析与决策。然而，不同模态数据的融合往往面临数据格式、分辨率、采样频率等异构性问题。为了让大模型有效地处理这些多样化的数据类型，需要开发统一的数据表示方法和高效的数据融合算法。这对提升大模型在多模态融合场景中的性能表现提出了巨大的技术挑战。

（2）行业知识与模型结合困难

制造业往往具有复杂的行业知识，涉及设备操作、工艺流程、质量控制等多个方面。如何将这些隐性和显性知识融入大模型，使其具备理解和应用这些知识的能力，是当前企业面临的主要技术难题。行业知识往往以非结构化的形式存在，因此需要通过知识图谱或本体论等方法进行结构化处理，使其便于与大模型整合。

（3）模型迁移与扩展困难

大模型在不同产品、工厂或场景中的迁移和扩展目前面临显著困难。不同产品的特性和工艺流程差异较大，直接迁移大模型可能导致性能下降。为了实现有效的大模型迁移，往往需要解决数据分布差异、特征不一致、标注数据不足等问题。少样本学习和领域适应技术是应对这些问题的潜在解决方案。

（4）模型输出的准确性差

在工业应用中，大模型的准确性将直接影响到生产效率和产品质量。模型输出的准确性受数据质量、模型选择、超参数设置等多个因素的影响。然而，工业环境中的数据往往存在噪声大且不完整的问题，导致模型的预测结果准确性低。为此，需要采用数据增强、模型选择优化、超参数调优等技术提高模型的输出准确性，从而确保大模型在实际应用中的可靠性。

2.3.2 大模型+工业知识

2.3.2.1 概述

大模型具备强大的语义理解能力，但是训练语料中的工业知识量较少，结合工业知识可以拓展大模型在工业领域的能力，目前业内对此的探索进展深入，以下是对“大模型+工业知识”应用方向的技术现状和技术挑战的详细介绍。

2.3.2.2 技术现状

大模型的工业知识的结合主要通过知识图谱增强、向量知识库增强、或其他检索增强的方式。知识图谱为大模型提供结构化的工业数据，使大模型更好地理解复杂的工艺流程和机制；向量知识库借助出色的语义检索能力为大模型提供外置的工业领域知识补充，同时增强大模型输出的可解释性和可溯源性；其他检索增强的方式例如Elasticsearch检索增强、搜索引擎检索增强等为大模型提供工业知识增强，提升大模型工业领域表现。

（1）知识图谱增强技术

知识图谱在工业场景中提供结构化的数据，用以帮助大模型更好地理解工业领域中的工艺流程、设备运作等复杂机制。通过将工业知识图谱与大模型结合，可以提升大模型在信息检索和推理任务中的表现。现有研究主要通过将结构化的工业知识转化为大模型可理解的输入形式，帮助大模型在理解复杂关系和推理时表现出更高的准确性和可解释性。例如，注入结构化知识图谱可以显著提高大模型在信息抽取和少样本推理场景中表现。

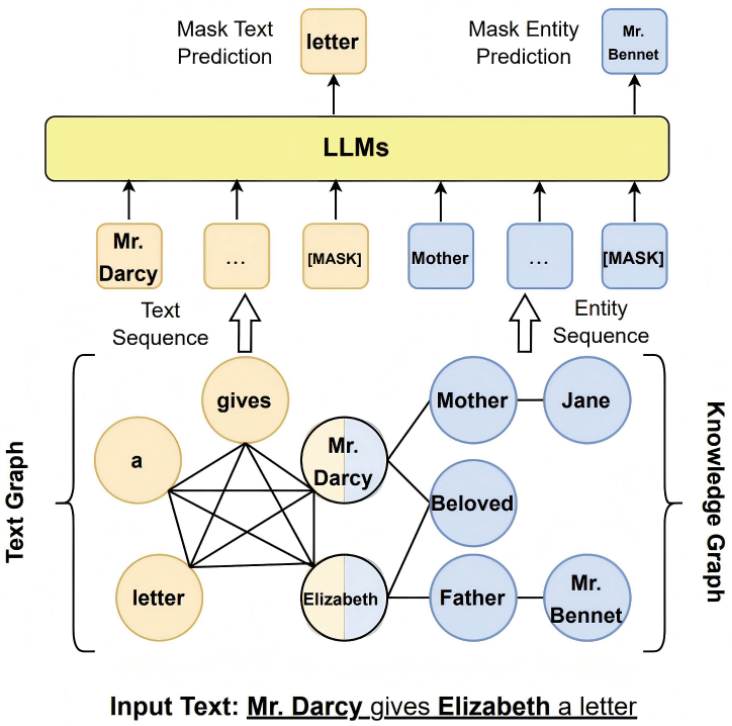

图2-10 知识图谱增强大模型技术示例[7]

（2）向量数据库增强技术

向量数据库作为大模型在检索增强中的关键组件，通过向量化技术存储和检索工业领域的语义信息，支持大模型从外部知识库中高效获取所需的工业知识。这类数据库以其高效的语义检索能力，提升了大模型的泛化能力和知识覆盖率。例如，在检索增强生成（RAG）框架下，向量数据库帮助大模型快速查找和生成相关的工业信息，增强了知识的可追溯性和输出的可靠性。

[7] 资料来源：Unifying Large Language Models and Knowledge Graphs: A Roadmap

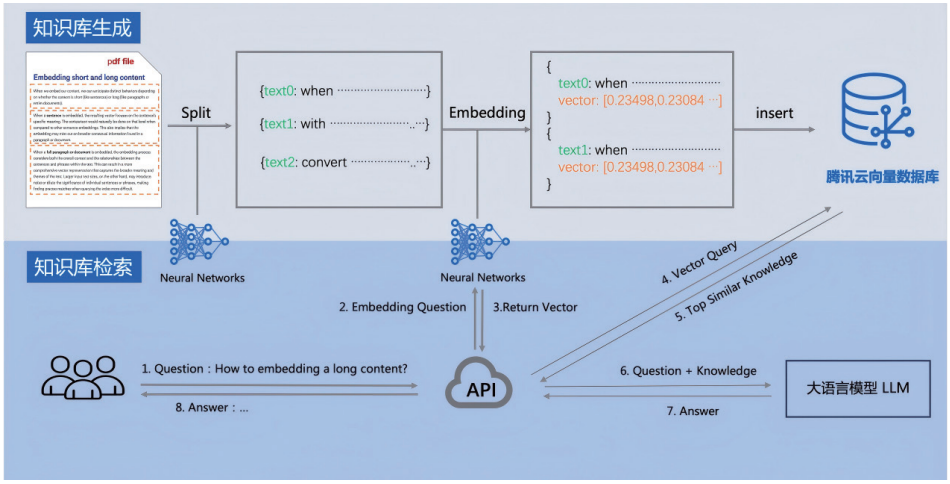

图2-11 大模型与向量数据库结合技术路径示例[8]

（3）综合检索增强技术

综合检索增强技术通过外部检索系统实时为大模型提供知识补充，从而提高其在工业领域中的应用表现。这类技术包括Elasticsearch等传统全文检索引擎，以及更先进的GraphRAG（图检索增强生成）技术。Elasticsearch通过快速检索和索引文档为大模型提供支持，而GraphRAG则结合知识图谱和生成模型，通过图结构数据的语义检索补充大模型的推理能力。这些技术能够动态调用外部知识源，提升大模型在复杂工业知识领域的准确性和可解释性。

[8] 资料来源：腾讯云官网

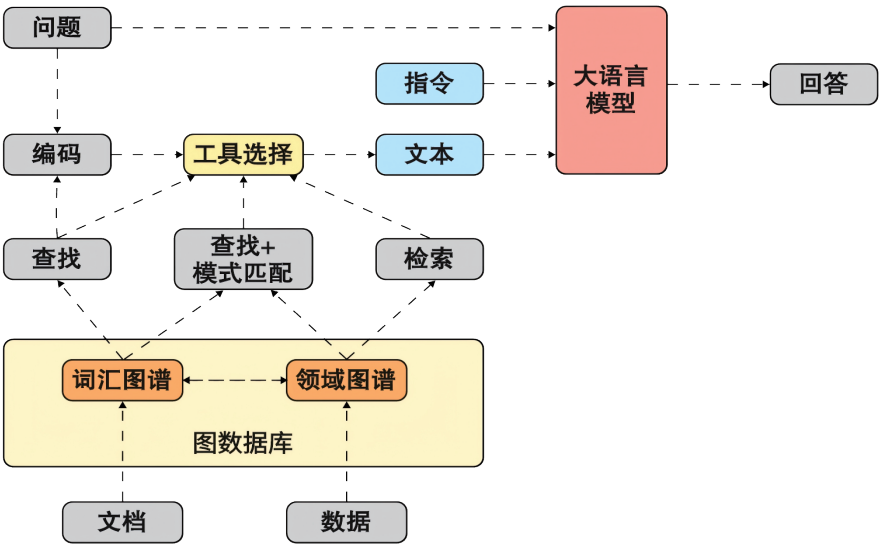

图2-12 综合检索增强技术示例GraphRAG技术路径[9]

2.3.2.3 技术挑战

（1）知识数据质量较差

大模型与工业知识结合的应用高度依赖于工业数据的质量。然而，工业数据常常存在噪声大、数据缺失、数据格式不一致等问题，可能会影响大模型在知识提取和推理中的表现。为了确保模型的可靠性和准确性，必须建立严格的数据清洗与预处理机制，以提升数据的一致性和完整性，确保大模型能够从中提取有用的知识。

（2）知识整合与表示复杂

来自多源异构数据的整合和知识表示是知识增强大模型面临的一个重大挑战。工业数据通常分布于不同的系统和格式中，包括结构化数据（如传感器数据）和非结构化数据（如文本和文档）。如何有效整合这些数据

[9] 资料来源：https://graphrag.com/

并通过知识图谱和其他技术实现统一表示和推理，依然存在技术难度，需要复杂的算法和模型设计。

（3）计算资源消耗巨大

知识增强大模型需要大量的计算资源，尤其是在处理大规模的知识图谱和复杂的推理任务时。训练和推理过程对计算资源的需求极高，这对企业的硬件设施提出了严苛的要求。因此，需要通过优化模型结构和合理配置硬件资源，来平衡计算需求与企业基础设施的所能提供的算力资源。

（4）安全性与隐私保护要求高

在处理敏感的工业数据时，知识增强大模型必须高度重视数据的安全性和隐私保护。工业数据通常涉及企业机密，任何数据泄露或未经授权的访问都会造成严重后果。因此，需建立严密的数据保护机制，确保数据存储、处理和输出的安全性，同时对输出结果的可解释性和透明性方面也要制定合规要求，以确保模型的信任度和可靠性。

2.3.3 大模型+小模型

2.3.3.1 概述

大模型与小模型的结合为工业应用提供了强大的计算能力与灵活的部署方式，尤其在需要复杂推理、大规模数据分析，以及对实时响应要求较高的场景中表现突出。大模型负责高计算需求的任务，而小模型则在边缘设备上实现快速、轻量的处理。这种组合在智能制造、自动驾驶、能源管理等工业场景中逐渐展现出重要应用价值。

2.3.3.2 技术现状

（1）模块化解耦对齐技术

模块化解耦对齐技术是大模型与小模型结合中的核心手段之一。通过将大模型与小模型分离为功能模块，并定义明确的接口进行数据交换与任务分配，可以在保持各自功能的同时实现灵活协作。大模型通常处理高复

杂度的任务，如全局优化与大规模数据推理，而小模型则专注于实时数据采集与边缘处理。模块化解耦使得两者之间的兼容性问题得以缓解，同时便于在不同计算资源环境下独立优化各自的任务。通过这种方式，大模型与小模型的协同工作变得更加高效且灵活，能够在不同工业场景下满足多样化的需求，提升系统整体的可扩展性与维护性。

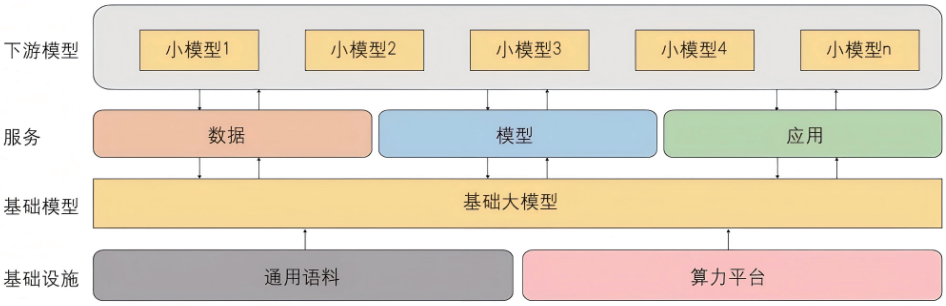

图2-13 大模型+小模型的数据与应用交互[10]

（2）边缘计算与云计算融合技术

大模型与小模型的融合通常涉及云计算与边缘计算的协同工作。在智能制造领域，大模型通常部署在云端，负责处理海量历史数据和复杂的分析任务，而小模型则部署在靠近生产设备的边缘节点，用于实时数据采集与初步处理，从而降低数据传输延迟并提升响应效率。例如，在智能生产线上，小模型可以实时监测传感器数据，如设备振动、温度等，快速识别异常状况以保障生产安全，而大模型则通过对历史数据和多工厂运营数据的深度分析，提供优化的生产排程和设备维护策略，实现整体效能提升。

[10] 资料来源：《姚前：关于大模型生态建设的若干思考》

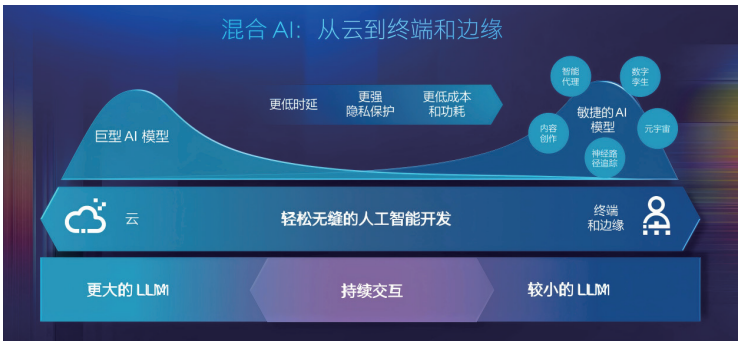

图2-14 边缘计算与云计算融合案例[11]

（3）联邦学习与分布式训练技术

联邦学习在大模型与小模型的结合中发挥了重要作用，特别是在数据隐私保护需求较高的工业场景。通过联邦学习，大模型可以从不同的边缘设备上学习模型参数，而无需直接访问敏感数据。这种分布式训练方式不仅提高了模型的泛化能力，还降低了数据传输和集中存储过程中的安全风险。

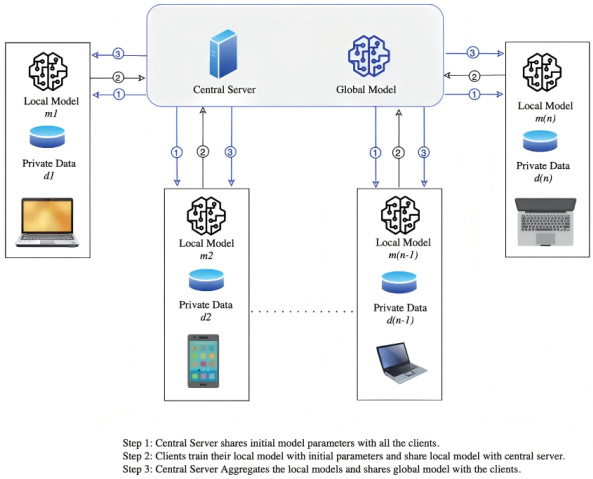

图2-15 联邦学习隐私保护模式示意图[12]

[11] 资料来源：英特尔官网

[12] 资料来源：A survey on security and privacy of federated learning

2.3.3.3 技术挑战

尽管大模型与小模型的结合带来了许多技术优势，但也面临以下重要的技术挑战：

（1）模型兼容性差

大模型与小模型在架构、数据格式和接口上可能存在较大差异，这在集成时面临兼容性问题。大模型通常使用复杂的算法和高计算量的架构，而小模型则倾向于轻量化，适合嵌入式系统或边缘设备。为了使两者有效协同工作，数据格式的一致性、接口的互操作性，以及模型之间的通信标准必须在设计阶段充分考虑。这不仅影响模型的精度，还关系到模型集成后的稳定性和响应速度。

（2）计算资源需求高

大模型的计算需求非常高，尤其是在处理复杂的推理任务时，需要部署在高性能计算集群或云端，对硬件资源的要求非常高。而边缘设备上的小模型虽然轻量化，但也需要具备足够的计算能力来处理实时数据。特别是在边缘计算场景中，需要有效协调云端和边缘计算资源，以确保大模型与小模型的协同运行不会对整体系统性能造成负面影响。

（3）实时性难以满足

工业应用场景对实时性和响应速度有非常高的要求，特别是在生产线、自动驾驶等需要即时决策的领域。由于大模型的复杂性和高计算需求，其推理时间可能较长，导致响应延迟。如何优化大模型的架构以减少计算延迟，确保小模型在实时场景中的快速响应，是实现工业场景中”大模型+小模型”有效应用的关键。

（4）安全性与隐私保护要求高

在大模型与小模型的结合过程中，安全性和隐私保护至关重要。由于工业数据通常涉及企业机密，任何数据泄露或未经授权的访问都会带来严重的后果。为此，必须采用严格的数据加密、访问控制和隐私保护技术。

联邦学习和差分隐私等技术可以有效确保模型训练和推理过程中的数据隐私安全，同时防止数据在传输过程中的泄露。

2.3.4 大模型+装备/应用

2.3.4.1 概述

大模型在装备应用中展现出强大的多模态数据处理和深度学习能力，可将文本、图像、传感器数据等多种信息融合分析，为设备运行提供精准预测与优化控制。同时，大模型通过自动化代码生成技术，简化设备编程流程并提升智能化水平。然而，其在兼容性和实时性方面仍面临挑战，需要优化模型设计与部署策略，确保其高效集成与实时响应，以更好地满足智能制造等复杂场景的需求。

2.3.4.2 技术现状

（1）多模态数据融合技术

大模型具备多模态数据处理能力，可以将不同类型的输入数据（如文

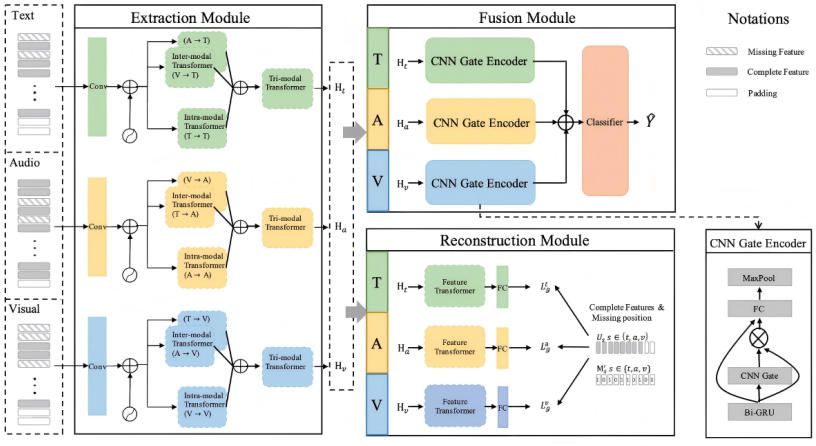

图2-16 多模态特征的融合与提取[13]

[13] 资料来源：Transformer-based Feature Reconstruction Network for Robust Multimodal Sentiment Analysis

本、图像、视频和传感器数据）进行融合处理。这一技术使得大模型能够从多个信息源提取相关信息，并进行综合分析，从而在工业装备的应用中提供更加精准的预测和控制建议。例如，在设备运行监控中，融合视觉和声音数据能够更好地检测设备故障。

（2）深度学习与设备优化

大模型通过融合深度学习算法，能够识别设备的运行模式，并通过优化控制参数提升设备性能。其中包括基于设备运行历史数据和实时数据的模式识别与预测，帮助优化工业装备的维护、操作和能效。设备通过大模型的智能分析，能够实现自我调节，最大化生产效率的同时减少故障率。

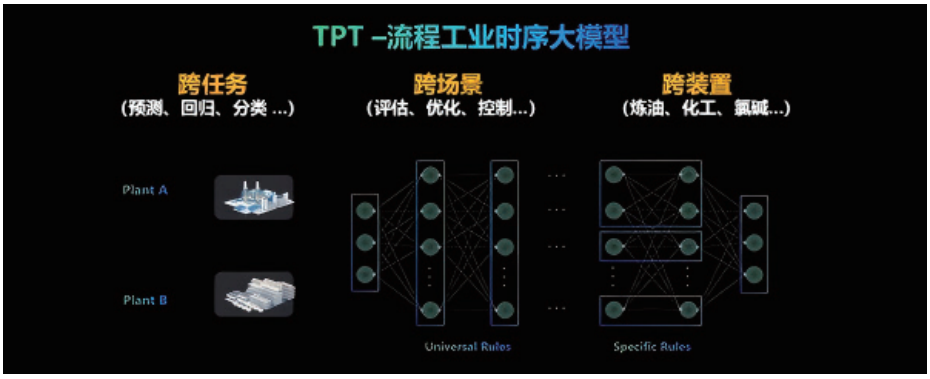

图2-17 时序大模型进行生产设备优化[14]

（3）自动化代码生成与设备编程

大模型可以通过自然语言处理和代码生成技术，为工业装备提供自动化编程支持。例如，基于设备需求描述，大模型可以生成或优化PLC（可编程逻辑控制器）代码，不仅简化了设备编程流程，还能通过机器学习算法提供编程错误预警和自动优化建议。

[14] 资料来源：中控技术官网

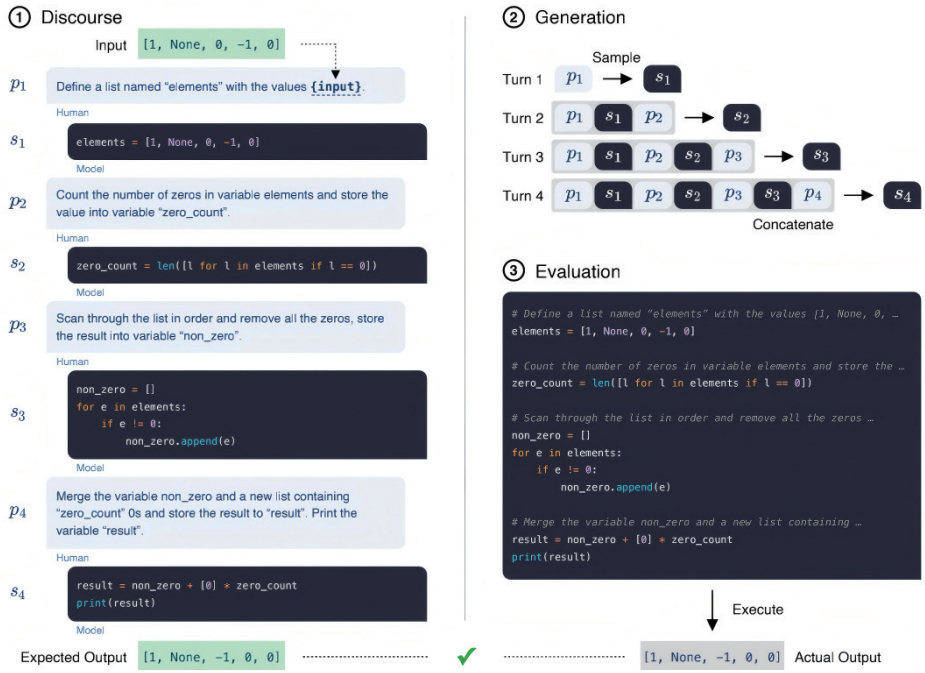

图2-18 大模型代码生成示例[15]

2.3.4.3 技术挑战

（1）设备与模型的兼容性差

在大模型与工业具体设备的集成中，硬件与模型的兼容性是一个关键问题。不同设备的计算资源、操作系统和数据传输方式可能不尽相同，如何确保大模型的运行环境与设备硬件的匹配，是模型设计和部署过程中需要克服的主要挑战。例如，一些工业设备的操作系统或硬件限制可能会限制大模型的部署和计算性能。

（2）实时性难以满足

工业设备通常需要对数据进行实时处理，以确保生产过程的连续性和效率。然而，大模型在处理复杂任务时计算量大，可能导致响应时间延

[15] 资料来源：A Conversational Paradigm for Program Synthesis

长，难以满足某些实时性要求高的场景。如何优化大模型的运行效率，减少计算延迟，是实现其在工业设备中有效应用的技术难点之一。

2.3.5 大模型+工具链

2.3.5.1 概述

在工业领域，大模型与工具链的结合为智能制造提供了强有力的技术支撑。多模态工具链整合文本、图像、传感器等多源数据，为复杂工业场景提供统一的处理与推理框架，实现高效的跨模态交互。通过标准化API与SDK的快速集成，企业能够大幅降低大模型应用门槛，加速智能化升级。同时，智能体技术依托工具链，实现自动化任务执行与实时决策优化。然而，工具链在多模型兼容性、响应速度及数据安全性等方面仍面临诸多挑战，亟需进一步优化以提升工业智能化水平。

2.3.5.2 技术现状

（1）多模态工具链技术

多模态工具链是大模型在处理多种数据类型（如文本、图像、视频和传感器数据）时的重要支持技术。它通过整合多种模态的数据源，提供了统一的数据处理与推理框架，使得大模型可以在不同场景中进行高效地交互与推理。例如，在机器人和人机协作领域，GitHub上的vlm_arm项目通过工具链，将多模态数据（图像、传感器数据等）与大模型结合，实现复杂的交互功能。这些工具链在处理多模态数据时，能够根据不同数据源的特性进行相应的处理，展现出极高的灵活性与应用潜力。

图2-19 苏秦语言大模型全工具链示意图[16]

（2）API与SDK的快速集成技术

随着大模型应用需求的增加，集成大模型变得越来越普遍。许多大模型平台通过提供标准化的API和SDK，简化了开发者在实际应用中对大

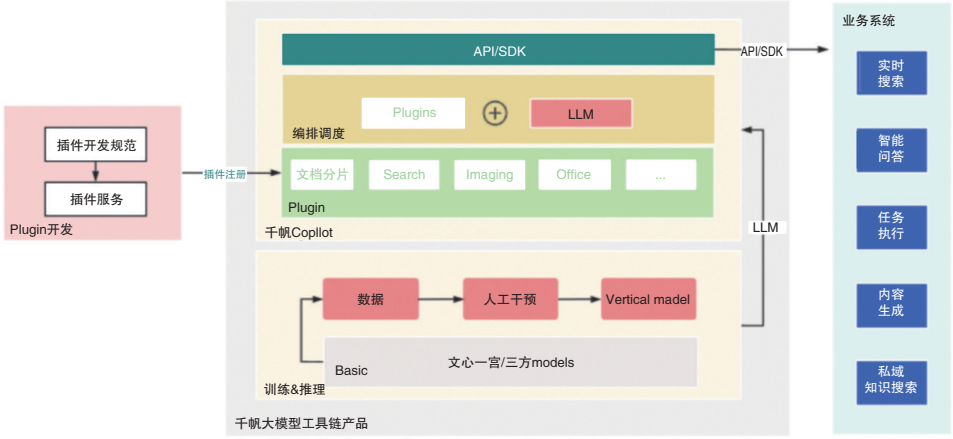

图2-20 百度千帆工具链API与SDK快速集成示意图[17]

[16] 资料来源：LLM-Kit 项目
[17] 资料来源：百度智能云

模型的集成。开发者只需利用这些平台提供的SDK和API，就可以轻松调用大模型的推理、训练和多模态处理功能，无需深入了解模型的底层架构。例如，百度的千帆平台提供了丰富的SDK工具，允许开发者快速集成ERNIE-Bot等大模型，进行文本处理、对话系统等多种任务。这种标准化的接口大大降低了大模型的开发门槛，使其在各种应用场景中得到普及。

（3）智能体（Agent）集成技术

智能体（Agent）是工具链中的一种应用技术，专门用于执行自动化任务并与外部环境进行交互。智能体利用大模型的推理与优化功能，能够执行自主任务，如环境感知、决策支持和自动化操作。通过工具链提供的API接口，智能体能够与大模型无缝协作，完成复杂任务。例如，在自动化生产或机器人系统中，智能体通过大模型工具链进行智能调度和任务执行，展现出高效的任务执行能力和适应性。

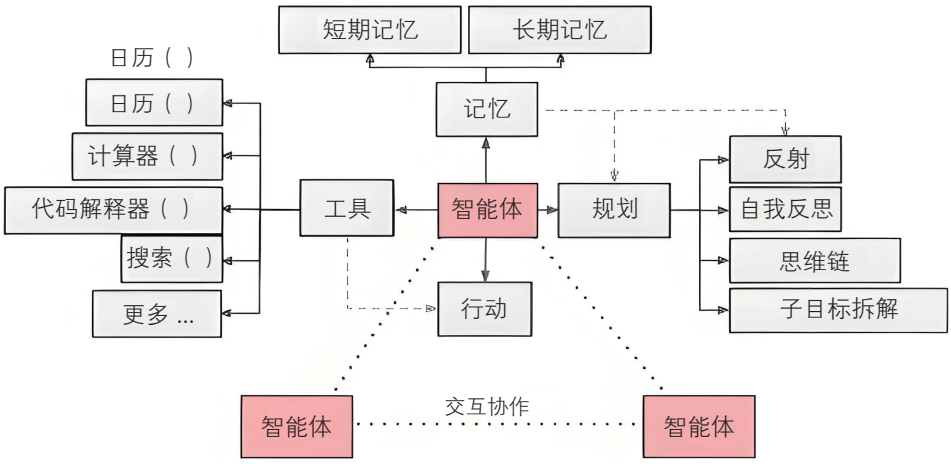

图2-21 大模型智能体集成技术[18]

[18] 资料来源：LLM Powered Autonomous Agents

多智能体系统是多个智能体组成的集合，它的目标是将大而复杂的系统建设成小的、彼此互相通信和协调和易于管理的系统。多智能体系统具有自主性、分布性、协调性，并具备自组织能力、学习能力和推理能力。在解决实际应用问题中，多智能体系统表现出较强的鲁棒性、可靠性和较高的问题求解效率。

图2-22 鼎捷indepth AI多智能体平台

2.3.5.3 技术挑战

（1）多模型兼容性较差

大模型平台和框架之间存在较大的接口差异，工具链在开发中面临多模型兼容性的问题。不同的大模型（如OpenAI、百度千帆等）拥有不同的API标准和接口设计，如何构建一个兼容多种大模型的工具链，是工具链开发中的复杂挑战。开发者需要设计统一的API接口，使工具链能够在不同模型之间实现无缝集成和切换。

（2）响应速度较慢

工具链需要在保持高效数据处理能力的同时，满足大模型实时性和低

延迟的要求。大模型在推理和处理海量数据时计算量较大，如何在这种高负载下保证实时响应，是一大技术挑战。尤其是在智能体（Agent）系统中，工具链必须能够在短时间内高效调度计算资源，以支持Agent的实时决策和任务执行。

（3）安全性与可控性要求高

工具链需要处理大量的敏感数据，包括企业运营数据、生产流程信息等，确保这些数据在传输和处理过程中的安全性尤为重要。此外，智能体在执行任务时可能涉及对敏感操作的控制，如何确保其行为的可控性，最大限度降低出现错误决策的风险，也是技术上的一大挑战。工具链必须提供完善的加密机制和权限管理功能，确保数据和操作的安全与合规。

第三章 面向智能制造的工业大模型参考架构

在面向智能制造的演进中，工业大模型的广泛应用正成为提高生产效率和产品质量的重要手段。本章将深入探讨工业大模型技术框架的设计和实施，这一框架不仅关乎技术的应用，更是标准化工作的核心基础。工业领域面临着海量数据的处理、复杂场景的分析、实时决策的制定等挑战，定义清晰的参考架构可以整合算力、存储、网络等基础设施，实现数据的高效处理与知识提炼，提供便捷的AI模型开发环境，并将AI能力转化为应用功能，形成一个完整、高效的智能化体系。部署架构是在技术架构的基础上，根据具体应用需求和资源状况进行设计，决定技术架构中各个层次和组件在实际环境中的部署位置和方式。技术架构和部署架构相互影响，相互优化。技术架构的改进可以推动部署架构的优化，例如通过模型压缩技术减少模型大小，从而降低对边端计算资源的要求；部署架构的实践经验也可以反馈到技术架构中，促进技术架构的进一步完善和优化。这些技术不仅有助于降低技术整合的复杂性，还能推动智能制造模型的广泛应用和互操作性，为我国智能制造的持续发展奠定坚实基础。

3.1 技术架构

工业大模型技术架构是指在工业领域应用大模型的系统设计和实现方式，这些架构需要存储和处理大量数据，具备强大算力，并且能够把工业领域的具体需求紧密结合，以实现智能化决策、自动化控制、质量检测、预测性维护等功能。工业大模型技术架构总体分为四个核心层次：基础设施层、数据层、模型层和应用层，每层承载着不同的功能与责任，共同支撑起整个工业智能化体系。

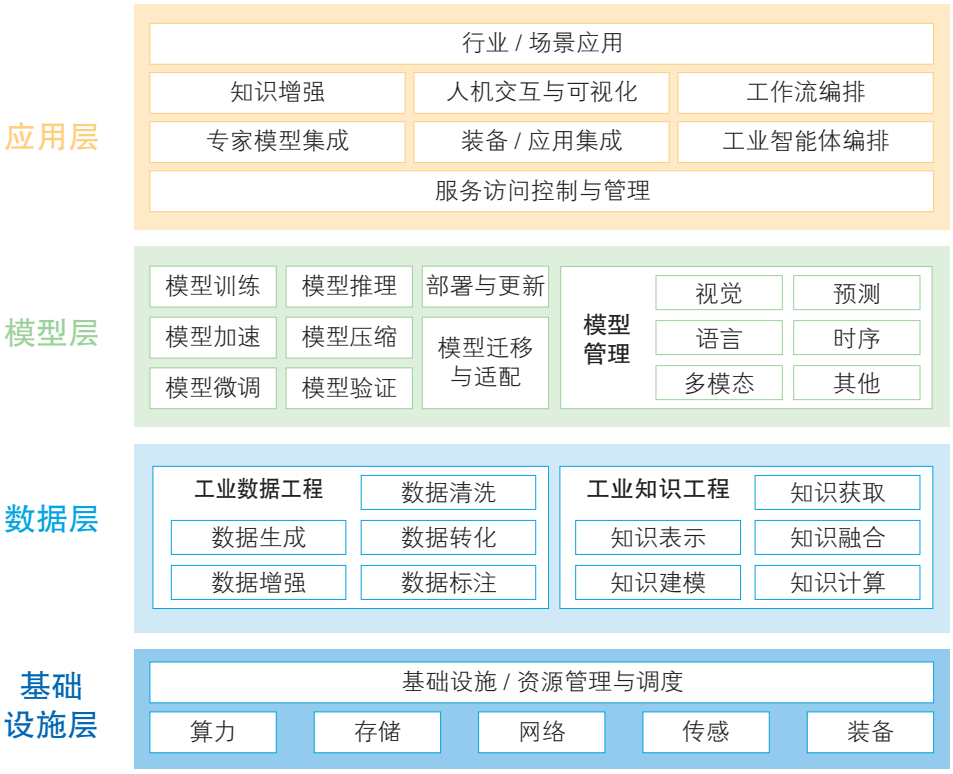

图3-1 工业大模型技术架构图

基础设施层：是工业大模型技术架构的基石，主要包括算力、存储、网络、传感器及装备等关键组件。算力涵盖了基于x86、ARM的通用计算，以及GPU、NPU、TPU等针对AI优化的计算资源，确保大模型具备强大的数据处理能力。网络通信包含对参数面、样本面、业务面（含存储业务面）数据的网络互通，主要采用系列化交换机，满足AI训练时对网络大带宽、低时延、高可靠等要求。存储采用分布式文件存储系统，支持智能分级，以优化成本效益。工业传感器作为工业物联网的关键组成部分，负责实时采集生产环境中的各种物理量（如温度、压力、位移等），并将这些数据转换为数字信号，供上层系统进行分析和处理。装备包括各种工业生产设备、机器人、自动化流水线等，它们与传感器、网络等基础设施

紧密结合，共同构成智能工厂的物理基础。这一层还包括统一的管理与运维，确保所有基础设施能够标准化、可交付且可持续运营。

数据层：不仅专注于数据的处理与知识的提炼，还承载着数据管理的重任，主要包含工业数据工程和工业知识工程。工业数据工程涉及数据生成、数据增强、数据清洗、数据转化及数据标注等关键技术，以确保数据的质量与可用性。工业知识工程通过知识表示与建模，将原始数据转化为有价值的知识资产，有助于将数据转化为可理解的、可应用的信息。为了实现这些复杂的数据处理与知识提炼任务，数据管理平台发挥着至关重要的作用，提供包括数据集成、数据开发、数据质量监控、数据目录梳理以及数据安全管理等功能，确保数据的准确性、一致性和安全性，从而促进知识的有效转化与应用。

模型层：是工业大模型技术架构中的核心部分，负责将数据层提供的原始数据和知识转化为可应用的AI模型，涵盖模型训练、模型推理、模型加速、模型压缩等一系列关键过程。随着应用场景的变化和需求的升级，还需要将模型迁移到新的平台或设备上，模型迁移与适配技术需要确保模型在不同平台和设备上的兼容性和性能表现。这一层还包含预训练基础大模型，通过在大规模通用数据集上进行预训练，学习语言的普遍特征和知识，具备泛化能力，如NLP、视觉大模型、多模态大模型、预测大模型、时序大模型等。在基础大模型上，利用行业数据进一步训练优化，可以形成行业大模型。针对特定应用场景，运用大模型工具服务套件，可以构建工业领域场景大模型。

应用层：是工业大模型技术架构的最终输出端，它将模型层的能力转化为具体的AI应用。这一层包括各种AI应用接口，如数据API、模型API、服务API等，以及多样化的AI场景化应用。在工业领域，应用层涵盖了知识增强、专家模型集成、装备/应用集成、人机交互与可视化、工作流编排、工业智能体编排以及服务访问控制与管理等功能，以全面支持

工业智能化的发展。

3.1.1 基础设施层

（1）高性能AI存储

高性能AI存储利用分布式架构和并行处理技术，将数据分散存储在多个节点上，并通过高效的网络连接实现数据的高速访问和处理。高性能AI存储的关键技术要点包括性能高密、大规模分布式部署、NFS/S3/POSIX多协议互通、数据编织技术、闪存存储、以数据为中心的架构、近存计算、KVCache以查换算、内置向量知识库和存储内生安全。这些技术可以实现全局数据可视可管、跨域调度，优化数据排布，减少AI训练周期，对数据集按需筛选、分级分类管理。

利用闪存技术的高速读写能力和低延迟特性，提高数据处理效率；采用高速互联总线（如CXL）将计算、存储、内存等资源解构为共享资源池，提升AI大模型的数据加载及流转效率，实现从以CPU为中心到以数据为中心的架构转变；通过算力卸载和随路计算，减少数据搬移带来的系统开销，提高数据处理速度，减少对GPU等计算资源的依赖；将知识内容转化为向量表示，实现多模态、高维度非结构化数据的快速查询检索；增强数据防护能力，通过使用数据加密和防勒索等技术，共同支撑起高效、安全的数据存储与处理框架。

在智能制造领域，高性能AI存储已经应用于多个场景。例如，在生产过程中，高性能AI存储可以帮助企业实现生产数据跨系统的快速归集和流动，优化生产流程；在质量检测环节，可以提供高性能的数据读写能力，支持实时质量检测和数据分析；在供应链管理中，可以实现跨域跨系统的数据调度和共享，提高供应链协同效率。此外，高性能AI存储还可以支持智能制造领域中的AI训练、模型推理等任务，推动智能制造向更高水平发展。

（2）AI大模型网络

AI大模型网络是实现数据中心内数据传输的重要通道，也是推动数据中心AI算力服务能力升级、实现算力充分释放的关键，具有高效、均衡、低延迟和高可靠性等特点，以满足大规模数据处理和实时决策的需求；能够实现流量的精细管理和智能分配，确保数据在传输过程中的高效性和稳定性。

工业大模型网络的关键技术要点包括网络级负载均衡（NSLB）、标准以太RoCE、数字平面快速恢复（DPFR）以及智能运维。这些技术分别针对AI训练场景下的流量不均衡、网络协议开放性、故障快速恢复和网络故障定位等问题提供相应的解决方案。通过NSLB技术，可以实现生产数据的均衡传输，避免网络拥塞和丢包，提高生产效率；RoCE技术则使得智能制造系统能够更灵活地选择合适的网络方案，降低投资成本；DPFR技术可以实现远程通告和快速换路，达到亚毫秒级收敛速度，保证智能制造系统在链路故障时能够快速恢复，提高生产过程的连续性和稳定性；智能运维技术可以实现故障逐跳检测，能够分钟级定位故障点，减少生产中断时间，提高智能制造系统的整体可靠性。

3.1.2 数据层

（1）数据生成

数据生成是指通过人工智能技术，基于已有的数据集和算法模型，自动生成新的内容数据。这一过程依赖于大数据分析和深度学习技术，能够模拟特定领域内的内容创作。针对工业领域，数据生成可以解决数据稀缺性、多样性不足等问题，为工业设计、制造、优化等环节提供丰富的数据支持。

在智能制造领域，数据生成技术可以用于产品设计环节。设计师可以快速生成多种设计方案，包括产品的外观、结构、功能等方面的模拟数

据。这些数据不仅可以帮助设计师快速评估不同设计方案的优劣，还可以为后续的制造和优化提供数据支持。除此之外，数据生成技术可以生成各种工艺参数下的仿真数据，包括切削速度、进给量、切削深度等，以便对制造工艺进行优化，不仅可以帮助工程师快速找到最佳的工艺参数组合，降低制造成本的同时提高产品质量。

（2）数据增强

数据增强是在已有数据的基础上，通过一系列技术手段（如图像变换、噪声添加、样本重组等）增加数据的多样性和复杂性，从而提高模型的泛化能力和鲁棒性。在工业领域，数据增强可以有效解决数据标注成本高、模型过拟合等问题，有助于提升工业智能系统的性能和稳定性。

例如，在自动驾驶的图像标注中，数据增强技术可以通过对图像进行随机裁剪、旋转、翻转等操作增加训练样本的数量和多样性，使模型可以在更多样化的数据上进行学习，提高泛化能力。同时，这些增强后的数据还可以作为预标注数据，供标注员进行确认或修改，进一步提高标注效率。在自动驾驶场景重现中，数据增强技术可以通过对真实场景数据进行变换和组合来生成新的训练样本，用以补充实际行驶过程中遇到各种复杂和罕见场景，如可以通过改变视角、改变光照、改变纹理材质的方法生成各种高真实感数据。这些增强后的数据不仅丰富了训练集的多样性，还有助于模型更好地适应不同环境和条件下的自动驾驶任务。

3.1.3 模型层

（1）工业大模型训练：MoE（混合专家模型）

MoE（混合专家模型）是一种基于Transformer架构的模型，它通过引入稀疏的MoE层和门控网络，实现在计算资源有限的情况下对更大规模模型或数据集的有效预训练。关键技术要点包括：稀疏MoE层替代传统的前馈网络层（FFN），MoE层包含若干“专家”，每个专家本身是一个独立

的神经网络；以及门控网络，负责根据输入数据的特征，动态地决定激活哪个专家模型以生成最佳预测。

在工业领域，MOE旨在解决大规模模型训练与推理中的效率问题，特别是在计算资源受限的情况下。它允许企业以更少的计算资源训练出性能更优的模型。在智能制造领域，MOE已应用于预测性维护和质量控制等场景，利用MOE模型对设备的运行数据进行实时分析，预测设备可能出现的故障，从而提前进行维护，减少停机时间和维修成本；通过MOE模型对生产过程中的各个环节进行监控和分析，及时发现并纠正可能影响产品质量的问题，确保产品质量的稳定性。

（2）工业大模型微调

工业大模型微调技术是指在已经预训练好的大型深度学习模型基础上，使用特定工业任务数据集对模型进行进一步的训练，使其适应特定工业任务或领域的过程。这种技术通过调整模型的参数和权重，使预训练模型能够针对具体的工业场景和问题提供更加精准和高效的解决方案。

工业大模型微调主要包括全参数微调和高效参数微调技术。全参数微调是对整个预训练模型的所有层和参数进行更新和优化，确保模型能够充分学习到工业任务的特定特征和规律。高效参数微调仅对模型的一小部分参数或引入额外的参数进行更新，如使用Adapter Tuning、Prefix Tuning或Prompt Tuning等方法，以减少训练负担并提高微调效率。

例如，在电子制造行业的微小电子元件装配检测任务中，通过针对该任务的微调，模型能够更迅速准确地识别元件的位置偏差、方向错误及焊接缺陷等问题，有效应对不同生产环境和产品变化带来的挑战，减少误检和漏检的概率，提升缺陷检测的整体效能。此外，工业大模型微调技术还可以应用于质量控制、预测维护、生产流程优化等多个智能制造领域的关键环节。

（3）工业大模型轻量化部署

轻量化部署是指将工业大模型进行压缩和优化后，部署到资源有限的设备上。轻量化部署能够降低大模型对硬件资源的需求，使其在资源有限的设备上高效运行，同时减少模型的复杂度和大小，加快推理速度，提高实时响应能力，并且有助于推动模型在移动端或嵌入式系统等场景中的广泛应用和普及。

轻量化部署包含模型压缩、分布式推理、硬件加速等关键技术。模型压缩是通过减少模型参数数量、降低计算量等方法来缩小模型的大小，常见的模型压缩方法如权重剪枝、模型量化、知识蒸馏、参数共享、低秩分解等。分布式推理可以将复杂的推理任务分解为多个子任务，并分配到多个计算节点上并行处理，提高模型的整体推理速度和效率。硬件加速利用针对特定任务设计的硬件加速器，如专用加速器NPU、图形处理加速器GPU、数据加密加速器、网络加速器等提高推理任务的执行效率。

例如，在产品质量控制环节，模型压缩技术使质量检测模型能够在生产现场的边缘设备上高效运行，自动进行缺陷检测，快速准确地识别出产品中的瑕疵或故障。在智能制造生产线上，通过安装传感器和摄像机捕捉生产数据，并利用分布式推理技术实时处理这些数据，对生产流程进行监控和优化。这些应用不仅提高了生产效率和产品质量，还降低了人力成本和时间成本。

在产品质量控制环节，采用专用的图像处理硬件，能够加速图像处理和分析的速度；模型压缩技术使得质量检测模型能够在生产现场的边缘设备上高效运行；利用分布式推理技术，可以实时处理这些数据。通过轻量化部署的大模型能够进行自动缺陷检测，以便快速准确地识别出产品中的瑕疵或故障，实现对生产流程进行监控和优化。

（4）工业多模态大模型

工业多模态大模型是将文本、图片、视频、音频等多种模态的信息联合起来训练而成的模型，不仅能够理解文字，还能解读图像、聆听语音，

甚至感知情感和动作，为工业应用提供了更加自然和丰富的人机交互方式。多模态大模型以训练好的自然语言大模型和编码器为基础，将编码器提取的图片等模态特征跟大语言模型的输入空间对齐，对齐后的图片等模态特征可以像文本特征一样输入到大语言模型中，使大语言模型可以支持图片等模态输入。

例如，在汽车造型设计领域，汽车造型概念设计草图的构思、多套风格方向效果图的设计、聚焦效果图的设计、定型效果图的设计等环节，往往会花费长达数月乃至上年的时间。基于汽车行业的工业多模态大模型能力，通过纯文字可以生成汽车内外饰草图、效果图，加上特定关键词，可以支持更多类型和场景化的效果图生成。工业大模型给汽车造型设计带来更多的创意与可能，有利于设计师持续发现新造型元素、探索更多未来概念，有效提升车辆造型的设计和创意效率，也可以帮助设计师快速筛选、比对、优化设计方案，减少无效设计，大幅加快产品迭代进度。

（5）工业时序大模型

工业时序大模型是一种针对工业时序数据进行建模和预测的人工智能模型。它利用先进的机器学习技术，对工业过程中产生的时间序列数据进行分析、建模和预测，以帮助企业实时监控和优化生产流程。工业时序大模型依赖深度学习算法进行建模，自动从数据中学习特征表示，构建非线性模型以捕捉工业时序数据的动态特性；运用频域多尺度学习算法在不同频率段上进行多尺度分析，捕捉数据中不同层次的信息；利用时序信息增强算法捕捉时间依赖关系，并结合机理知识与实际生产数据，通过专业知识描述工业过程，实现高精度、强泛化性的模型构建与优化。

例如，在设备健康状态监测中的场景中，传统的设备健康状态监测通过人工定期巡检实现，存在时效性不足、部分高空、密封空间作业困难等问题。通过工业时序大模型与自然语言大模型的结合，可以实现设备健康状态的24小时监测及分析。这一技术路线利用时间序列预测算法处理历史

能耗数据、天气信息等，准确预测未来能源需求，并通过机器学习算法进行能耗数据分析、故障检测与预防，不仅克服了传统监测方式的时效性不足及高空、密封空间作业困难等问题，还有效提升了MTBF（平均无故障时间），实现了从事后维修到预防保全的转变，为工厂提供了科学的决策支持。

（6）工业预测大模型

工业预测大模型旨在通过预训练模型推荐和模型融合等技术，解决表格数据中的多种任务问题，包括分类、回归、异常检测以及时序预测。分类任务：以客户流失预测为例，模型通过输入客户的购买频率、平均交易额、最近购买时间等多个特征，能够准确预测客户是否会流失；回归任务：在产品成本预测中，模型能够根据当前的市场环境、生产条件以及企业内部的管理状况等因素，自动调整预测结果，从而为企业提供准确的成本预测信息异常检测任务：在工业设备维护中，模型通过监测设备的温度、振动频率、电流消耗等多个运行参数，以及设备的维护历史和工作环境等信息，能够及时发现并识别出数据中的异常点，从而预防潜在的设备故障；时间序列预测：在电力负荷预测中，模型综合考虑过去多个时间段的电力消耗量、当前时间、日期、季节、天气条件等多个因素，能够准确预测未来N小时或N天的电力需求，为电力调度和规划提供重要依据。

（7）工业代码大模型

工业代码大模型是专门应用于工业软件开发和编程领域的大型语言模型，能够处理和理解大量的代码数据，旨在减少开发人员的手动编码工作，提升软件开发的效率和质量，帮助开发人员识别和修复代码中的潜在问题，提高软件的稳定性和可靠性，减少软件维护的工作量，降低维护成本。

工业代码大模型的关键技术包括：深度学习与自然语言处理，利用深度神经网络和自然语言处理技术，理解代码的语义和结构，实现代码的智能生成和修改；大规模参数和复杂计算结构，能够分析和处理大量的代码

数据，捕捉代码中的模式和规律；代码理解与生成能力，能够理解现有代码的逻辑和功能，并根据开发者的需求生成新的代码段；跨语言和多模态支持，支持多种编程语言，并能够处理代码、注释、文档等多种模态的信息，提供更全面的开发支持。

例如，代码生成，自动生成符合规范和功能要求的代码段；代码补全，为开发人员提供智能的代码补全建议，帮助开发人员快速编写和修正代码；代码编译优化，通过分析和优化代码，提高编译效率和运行性能，降低资源消耗；代码质量评估，对现有代码进行全面的质量评估，包括代码可读性、可维护性、安全性等方面；代码修改建议，针对代码中的潜在问题或改进点，提供智能的修改建议，帮助开发人员优化代码结构和性能；测试用例生成，根据代码的功能和逻辑，自动生成相应的测试用例，提高测试的覆盖率和效率。

（8）类脑大模型

类脑大模型旨在通过优化计算和数据传输能力，实现对复杂任务的高效处理。其实时监控、动态调整和故障修复功能能解决传统技术难题。它在设计、性能、稳定性和适应性方面需达到极高标准，以满足工业应用需求。类脑大模型包括以下关键技术：1）突触可塑性模拟：通过Hebbian学习规则和时序依赖性可塑性（STDP），类脑大模型能够动态调整神经元连接，实现自适应学习，逐步增强记忆、学习和决策能力；2）生物神经网络模拟：通过使用高精度神经元模型（如Hodgkin-Huxley和Izhikevich模型），类脑大模型能够准确模拟神经细胞电活动，真实再现大脑的信息处理方式，并动态优化网络结构；3）脉冲神经网络（SNN）：通过模拟大脑的脉冲式信号传输，神经元仅在信号达到阈值时激活。这种事件驱动方式提高了信息传递效率，尤其适用于低功耗和实时处理场景。4）时序信息处理：类脑大模型通过时序依赖性突触可塑性等机制，依据信号传输时间调整神经元权重，模拟大脑处理时序信息的优势，尤其在感知、决策和

记忆中表现突出。5）能量高效计算：类脑大模型借助SNN和异步处理机制，实现低功耗的高效计算。结合类脑芯片，模型进一步提高了能效，接近人脑的计算效率。

例如，在制造领域，它能优化生产线布局，提升设备故障预测和维护能力；在物流领域，它优化运输路径和仓储管理，提高运输效率的同时降低管理成本。

3.1.4 应用层

（1）云边端协同

云边端协同计算技术是一种分布式计算范式，借助云数据中心、边缘服务器和终端设备的分布式算力，提供一个高效、灵活和可扩展的计算框架。在这种模式下，云服务器、边缘服务器和终端设备可以在不同的计算层级之间进行协作，共同承担计算任务，从而显著提高整体的计算效率。

云边端协同训练和推理中，端侧（如物联网设备、智能手机等）负责实时收集各种传感器数据、用户交互数据，对于具有一定计算能力的端侧设备可以进行初步的数据处理、实时反馈模型推理结果。边侧（如边缘服务器）通常部署在网络边缘，靠近端侧，能够快速处理来自端侧的数据，减少数据传输延迟。边侧可以对来自多个端侧的数据进行预处理、聚合，再上传到云端，有助于减轻云端的计算压力。同时可以缓存常用的模型和数据，以便后续快速分发给需要的端设备或进行模型推理，满足实时性要求高的应用场景。云侧不仅提供强大的计算资源以支持大模型训练，还可以对训练好的模型进行深入评估和优化，以提升模型性能。此外，云侧还负责整个云边端系统的全局调度与管理，包括合理分配计算任务、有效协调资源，从而确保整个系统的稳定运行和高效协作。

例如，在智慧工厂的建设中，端侧通过边缘网关等载体与工业设备联接，实时采集设备及生产数据，同时，工业相机等数据也通过标准协议接

入。边侧通过工业物联平台可以实现设备互联和数据全面采集，对采集的数据进行初步处理，快速反馈推理结果，以支持即时的工业决策与控制。云侧以云平台为基础设施，基于工业物联平台（中心侧）、工业AI质检平台、工业大数据平台、云高阶服务、视频管理服务、存储服务等技术平台及工业AI大模型，提供工厂级生产业务应用及管理系统。

（2）工业知识增强

工业知识增强通过引入和融合外部知识库，对现有知识进行增加、删除、修改等操作，以提高工业知识的质量和深度，进而提高工业生产效率和产品质量，旨在将知识图谱、专家系统等知识工程技术与工业领域的实际需求相结合，使工业生产更加智能化和自动化。

工业知识增强的关键技术要点包括知识图谱的构建和应用、自然语言处理、知识推理与决策支持等。知识图谱是工业知识增强的基础，通过结构化的方式表示工业领域中的实体、属性以及它们之间的关系，以便从海量的工业数据中提取出有用的知识。在工业环境中，大量的信息是以自然语言的形式存在的，如技术文档、操作手册、维修记录等。NLP技术能够使得计算机系统理解和处理这些自然语言文本，从而提取出关键信息。知识推理技术能够基于已有的知识推导出新的结论，而决策支持系统则能够根据这些结论为工业生产过程提供智能化的建议和指导。

例如，在智能制造过程中，系统可以利用知识图谱中的工艺知识、设备知识等，对生产流程进行优化和调度，提高生产效率。同时，通过自然语言处理技术，系统可以理解和解释生产过程中的各种指令和反馈，实现更加精准的控制。此外，机器学习算法的应用也使得系统能够不断学习和改进，以适应不断变化的生产需求。

（3）科学计算

科学计算是利用大规模计算模型和复杂算法，针对自然科学、工程科学及社会科学等领域的复杂问题进行求解和分析的过程。其核心在于通

过高性能计算和先进的数据建模技术，从海量数据中挖掘数理规律，以高效、准确的方式解决科学研究和工程实践中的难题。

科学计算的关键技术要点包括高性能并行计算和数值模拟与流程仿真。高性能并行计算利用集群计算、GPU加速和分布式训练等技术，处理大规模数据集和复杂计算任务，提高计算效率和吞吐量。数值模拟与流程仿真技术则通过虚拟环境中的流程再造分析，减少实际试验成本和时间，优化设计和预测性能。

例如，在生物医药制造场景中，科学计算利用高性能计算和复杂算法，对海量的生物数据、化学结构信息、医药典籍、文献、临床数据等进行深度分析和模拟，加速药物分子的设计、筛选和优化过程，从而缩短药物研发周期，提高药物研发的效率和成功率，为新药的开发提供强有力的支持。

（4）具身智能

具身智能是指智能系统或机器通过感知和交互不断与周围环境进行实时互动，从而不断优化和调整自身行为的能力。它强调智能不仅仅依赖于算法和数据处理，更在于智能体与环境之间的互动和适应。

具身智能的关键技术要点主要包括感知与规划融合和具身执行大模型。感知与规划融合能够整合来自不同传感器的数据，形成对环境的全面感知，并基于这些信息实时进行任务规划和决策，制定出最优行动方案。它内置了丰富的领域知识和先验知识，有助于智能体做出更合理准确的决策。具身执行大模型关注于智能体在实际环境中执行复杂任务的能力，不仅能够理解语言指令，还能生成并指导智能体在真实环境中执行任务的指令，通过感知、规划和执行动作的循环，完成各种任务。

例如，在智能抓取分拣场景中，具身智能系统通过先进的视觉系统和算法，能够迅速识别并精确抓取不同形状、位姿的物料，自主规划抓取路径，避免碰撞事件发生，提高分拣效率和准确性。此外，在智能人机交互

领域，具身智能系统使机器能够深刻理解人类自然语言，并根据指令生成相应代码，实现更自然、更智能的人机交流，从而优化生产流程，提高生产效率。

（5）工业智能体

工业智能体是指在工业生产环境中，具备感知、决策、执行和学习能力的高度集成化智能系统，融合了人工智能、物联网、大数据等先进技术，旨在实现工业生产过程的智能化、自动化和高效化。

工业智能体的关键技术要点包括多模态感知与融合、智能决策与规划、精准执行与控制以及持续学习与优化。通过多模态感知技术，实时获取生产环境中的各种信息，如视觉、听觉、触觉等，并进行融合处理，以形成对生产环境的全面理解。基于这些信息，智能决策与规划技术能够制定出最优的生产计划和调度方案，确保生产过程的顺利进行。精准执行与控制技术负责将决策结果转化为具体的操作指令，并控制生产设备按照指令进行精确操作。持续学习与优化技术使工业智能体能够不断从生产过程中学习新知识，优化自身性能，以适应不断变化的生产需求。

例如，在智能工厂中，工业智能体可以通过收集生产线上的各种传感器数据，实时监控生产设备的运行状态和产品质量，及时发现并处理异常情况，确保生产过程的稳定性和可靠性。它也可以根据生产计划和需求，自动调整生产设备的参数和工艺流程，实现生产过程的灵活性和高效性。在智能物流、智能仓储等领域，通过精准感知和决策，实现物料的自动搬运、存储和配送，提高物流效率和可靠性。

3.2 部署架构

3.2.1 四级架构

该部署架构适用于集团/超大型企业或有分支机构对数据有管制要求

的企业，应用场景涵盖工业制造的所有领域，如图3-2所示，涉及两级训练和三级推理。

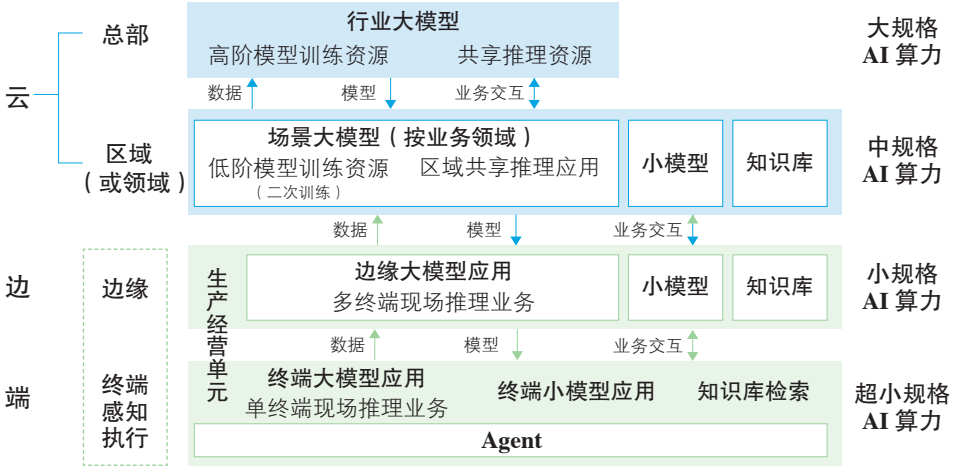

图3-2 工业大模型四级部署架构

对于模型训练，企业总部可根据自身能力选择研发与采购行业大模型，以构建集团内部统一共享的基础模型。区域（或领域）公司根据自身特点基于集团的模型进行二次训练，形成更符合实际业务需求的场景大模型。同时，可引入行业专属的小模型和知识库，用于补充大模型在某些特定场景下推理精度无法满足要求的场景。

对于模型推理，企业总部推理服务于全公司共享业务，包含为全公司服务的推理业务和为全公司服务的推理资源池，但不对有数据管制要求的公司分支机构提供推理业务；区域（或领域）推理服务于区域（或领域）业务，可将大模型、小模型和知识库集成到一个统一的框架中，支持多种推理的组合应用。

边端侧有大量的实时控制业务，需就近部署推理算法以满足业务响应时间，并在现场需适配不同的周边系统和本地化业务流程，便于集成适配和减少联动更新。在终端感知和执行过程中，智能体Agent可依据任务

复杂度和实时性需求，灵活决策使用大模型、小模型或检索的方式，以实现精度、效率、快速响应等方面的平衡，从而实现智能、高效的决策与执行。

针对某些复杂的应用，在执行推理时，可根据业务的逻辑设计通过云、边、端的推理资源协同完成。通过二次训练得到的细分场景大模型，应根据实际边端侧计算资源的大小、业务精度要求进行合理压缩。

3.2.2 三级架构

该部署架构适用于大中型企业，应用场景涵盖工业制造的大部分业务领域，如图3-3所示，涉及一级训练和二级推理。部署架构可进一步简化，在模型训练与微调方面，聚焦构建企业内统一共享的场景大模型或行业大模型，以便在此基础上为工厂、车间、设备提供推理服务。企业可根据自身能力选择开发或从外部直接获取行业大模型及其能力，并根据自身特点进一步微调训练，形成更符合实际业务需求的场景大模型。同时，通过引入或构建专属的小模型和知识库，用于补充大模型在某些特定场景下推理精度无法满足要求的场景。

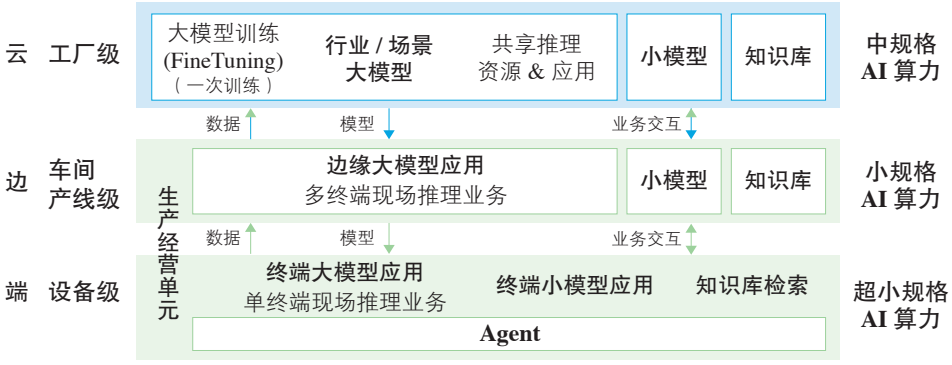

图3-3 工业大模型部署三级架构

3.2.3 二级架构

该部署架构适用于中小型企业，应用场景涵盖工业制造的部分业务领域，如图3-4所示。与企业规模相适配，部署层级可进一步调整为两级，覆盖车间级和设备级，并降低对算力设施的需求。与云服务器相比，终端设备的计算能力、存储空间和内存可能有限，持续的计算任务可能会增加端侧设备的功耗。需通过模型压缩、蒸馏等技术减小模型规模，以保障模型的流畅运行。另外，需考虑操作系统和端侧设备对模型的兼容性，以及大量部署端侧模型的设备管理问题，以保障模型在设备部署数量的不断提升。

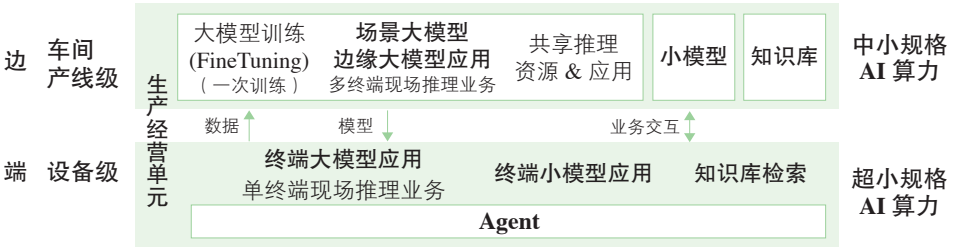

图3-4 工业大模型部署二级架构

3.2.4 企业获取和构建基础大模型能力的途径

企业获取和构建基础大模型能力的途径有以下方式：

（1）采用云服务平台：许多云服务提供商提供了强大的AI平台和工具，企业可以通过订阅这些平台的服务，获得大模型能力。这种方式不需要自己构筑底层算力平台，省去复杂的运营和维护工作的成本，大模型开发可以做到即开即用；这种方式可以节省企业的研发成本和时间，同时还能享受云服务提供商的技术支持和稳定性。

（2）合作开发：企业可以与专业的AI公司或研究机构合作，共同开发和训练大模型，这类专业公司和机构可以提供技术支持、算力资源、数

据等方面的帮助，加速大模型的研发和应用。此类方式需要企业和专业公司一起构建底层算力平台，并承担日常运营维护管理。

（3）自行开发：企业可以自行组建研发团队，进行大模型的研发和训练。该方式需要企业具备强大的技术实力和资源投入，包括算力、数据、人才等方面的支持。

第四章 应用场景

从产品设计、生产、物流、、服务到运营，工业大模型的应用正不断重塑着工业产品的全生命周期。本章节旨在提供一个全面的视角，以探讨和展现工业大模型在工业全流程中的应用。本章节总体架构如图4-1所示。

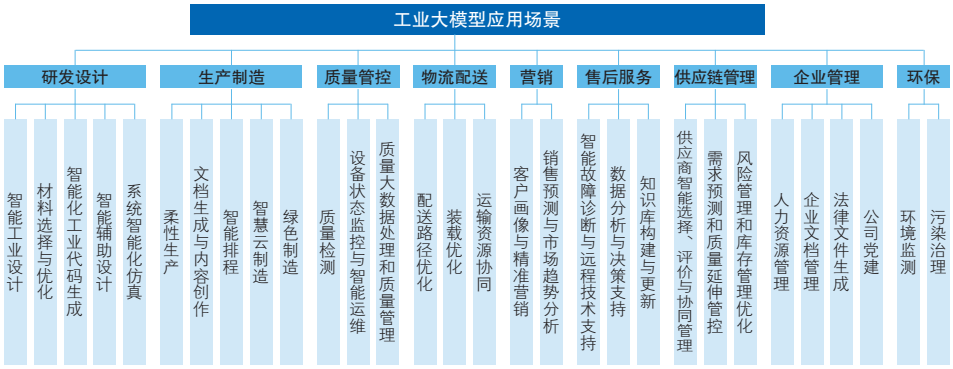

图4-1 应用场景总体架构

4.1 研发设计

4.1.1 智能辅助设计

大模型通过整合跨学科知识，如机械工程、材料科学、电子技术等，对海量设计案例、技术文献以及专利数据进行学习，为研发人员提供创新性的设计思路和概念。通过输入关键词或描述，让大模型生成初步的设计创意或文案；利用GANs、Diffusion Models生成高质量的产品外观图像、音频等内容；利用多模态大模型同时处理文本、图像、音频等多模态数据，实现跨模态的信息融合和理解；结合多维度设计细节，实时优化大模型，快速处理输入数据并生成实时反馈，支持设计师在设计过程中的即时调整和优化，最终生成工业设计方案。

例如，海尔集团利用工业大模型技术加速家电产品设计创新，通过生成的设计效果图和方案，设计师、工程师、市场专家和客户服务团队能够

提前介入设计流程，提供专业反馈和市场洞察。这种早期的、多角度的协作有助于在产品设计阶段提前发现潜在问题，减少后期修改成本和时间，进而加快产品上市速度。

又例如，华天软件在PLM系统中开发了“华小天”PLM智能助手，通过深度学习和AI大模型技术，实现对PLM结构化、非结构化数据智能检索，智能推荐等功能，显著提升企业研发设计知识重用率和产品数据准确率，减少重复劳动与返工，从而提高整体研发设计效率；在CAD系统中，通过AI大模型与云CAD（CrownCAD）开发框架的深度融合，实现语义翻译引擎、智能流程自动化等功能，提供基于三维智能快速检索、AI辅助建模设计等智能化应用，进一步提高研发设计效率。

4.1.2 系统智能化仿真

通过大模型技术，创建产品的虚拟原型，并在虚拟环境中进行测试。

利用大语言模型的广泛知识库和上下文理解能力，准确捕捉用户需求，生成更符合用户需求的测试方案，从而辅助用户设置虚拟原型测试的参数、场景等。

结合计算机辅助工程（CAE）等工业软件，如有限元分析（FEA）、计算流体力学（CFD）等，对产品在各种复杂工况下的性能进行全面仿真，提前发现潜在的设计缺陷和性能瓶颈，并进行针对性的优化改进，减少物理样机的试验次数和研发成本，提高产品的可靠性和耐久性。

根据不同模态信息，提供更全面的测试视角和反馈，在测试过程中动态注入故障场景，模拟实际运行中的异常情况，以评估虚拟原型的鲁棒性和稳定性。

4.1.3 材料选择与优化

利用深度学习和机器学习算法对材料性质、结构、性能之间的联系进行学习，建立对应模型，针对不同材料的性能、成本和可持续性进行评

估，分析材料的微观结构、化学成分与宏观性能之间的关系，实现从原子尺度到宏观尺度的材料全面设计和预测，通过改变拓扑结构来增强材料力学性能和改善功能，并指导材料的合成与制备工艺优化流程，为产品设计提供最优的材料选择建议，实现产品性能的最大化；构建数据驱动的材料选择与优化平台，实现材料信息的快速检索、分析和优化。

4.1.4 智能化工业代码生成

应用于工业代码生成，支持多种编程语言和平台，根据工程师的自然语言描述，严格按照工业编程的最佳实践和标准，自动生成符合相关工业标准的代码，并可快速生成准确无误、重复性高、逻辑简单的代码，使得从设计到生产的转换更加迅速和流畅，提供针对不同工业应用场景的代码生成解决方案。

4.1.5 智能工艺设计

构建工艺知识图谱工艺自动推理、基于三维模型搜索自动匹配的工艺快速创建以及工艺知识智能问答应用。

基于图数据库语义的表达和管理，构建并利用工艺知识图谱，进行知识发现，知识挖掘和利用，实现工艺设计的知识匹配及推送。

基于零件自动匹配的工艺快速创建，采用大模型和AI深度算法，通过对不同三维模型的解析和对数据的预处理，提取模型的特征，实现三维模型的分类和入库。进行三维模型搜索时，通过特征匹配输出相似模型和关联的结构化数据。与三维搜索技术实现对接，支持数据聚类以及查询相似零组件。

例如，华天软件创新性研发出基于“知识图谱耦合小样本学习”的智能工艺生成应用，即SVMAN工艺系统。该解决方案通过构建多维工艺知识图谱，实现领域知识结构化建模，涵盖制造目标参数、工艺特征智能、特征工艺方法链、材料特性约束等关键维度，并通过建立工艺知识语义网

络和知识蒸馏，将领域专家经验与模型深度融合；同时开发迁移增强型小样本学习算法，在训练数据量较少的情况下仍保持90%以上的工艺推理准确率。该解决方案通过知识工程与深度学习的深度融合，开创了工艺设计从"经验驱动"向"知识-数据双驱动"的范式变革，为制造业智能化转型提供创新引擎。

4.2 生产制造

4.2.1 柔性生产

工业大模型在柔性生产领域的应用主要体现在生产线优化、设备预测性维护等方面。

对生产现场进行实时监控。实现对视频流中目标的识别、跟踪和定位，及时发现异常情况，如设备故障、人员违规行为等，并分析员工的操作行为，预测是否存在违规操作风险。此外，可以自动调整生产参数，实现生产过程的自适应控制；利用历史运行数据，分析模型预测并计算设备故障的发生时间和概率，结合设备维护手册和专家知识库，提供详细的维修指导和解决方案。

针对生产作业中的资源分配、任务调度等问题，对模型进行优化，使其能够在满足约束条件的前提下，寻找最优的调度方案，提高生产效率和资源利用率，并且通过动作级的工艺机理建模，实现装配任务序列及最优控制参数的精准规划，生成机器人控制代码，最大限度缩短工艺设计周期。

例如，华天软件的MOM系统，通过质量历史数据训练模型，预测工艺参数对质量的影响，实现智能质量预测与缺陷检测，同时可基于订单优先级、设备状态、物料库存，实时生成最优排产方案，还具备工艺参数智能推荐、预测性设备维护等功能。该系统将AI大模型与MOM平台进行深度融合，显著提高生产效率和产品质量。

4.2.2 文档生成与内容创作

（1）技术手册与用户指南的自动撰写

通过自然语言处理技术，自动从技术参数和产品功能描述中提取关键信息，生成详尽且易于理解的技术手册和用户指南。同时确保信息的准确性和一致性，为用户提供清晰的操作指导。

（2）维护文档与故障排除向导的生成

针对设备的维护和故障排除，根据历史维护数据和常见问题，自动编制维护流程和故障处理指南。帮助维护人员快速定位问题并采取相应措施，减少设备停机时间，提高生产效率。

（3）生产与市场分析报告自动化编制

利用自然语言处理技术，从生产数据和市场动态中提取信息，自动生成详细的分析报告。提供数据支持，帮助工程师做出最优决策，同时也为市场策略调整提供依据。

4.2.3 智能排程

构建企业资源管理系统，结合工业大模型能力，实现生产计划的自动化和智能化优化。实时监控和采集生产线上设备运行状态、原材料库存水平、人员排班情况以及订单交付信息等数据，应用运筹学等算法，制定出最优生产计划和排程方案。例如，在电子制造企业中，针对电子产品的不同型号、配置和订单需求，工业大模型可根据设备产能、物料供应、工艺路线等约束条件，合理安排生产任务的优先级和生产顺序，实现生产线的高效切换和资源的优化配置。

4.2.4 智慧云制造

云制造的概念最早于2009年由李伯虎院士提出，工业大模型将更大程度地贯彻云制造“制造即服务”的理念，实现更加智能化的云制造流程：

（1）需求端图纸的智能化解析和智能化自动报价

基于计算几何、3D几何模型引擎及3D模型分析，实现需求端加工图纸的自动解析，并基于加工工艺、材质、加工精度等参数实现快速智能化报价，实现加工询价的快速反馈和订单达成。

（2）海量订单与海量协同制造资源的智能化快速匹配和调度

在海量订单的加工需求中匹配最合适的加工资源，并实现快速加工任务分发和快速上机。

（3）制造加工可行性的智能化分析

无需人工判断图纸设计问题，降低试错成本，实时输出反馈。

（4）虚拟化与服务化

将各类制造资源和制造能力虚拟化、服务化，在工业大模型中构成制造资源和制造能力的云服务池，用户可以根据云池快速找到符合自身需求的制造资源与服务。

4.2.5 绿色制造

工业大模型形成云制造模式的资源共享、信息共享和协同创新，提高了制造效率和质量，降低了成本和风险：

（1）减少闲置产能

通过共享制造资源，如生产设备、专用工具、生产线等，提高生产资源的利用效率，减少了闲置产能。

（2）降低企业对自投自建制造资源的依赖

通过将分散、闲置的生产资源弹性匹配、动态共享给需求方，扩大云制造的有效供给，降低了企业对自有制造资源的依赖。

（3）优化制造成本

通过调动企业各类制造资源进行协同合作，以及优化各类资源配置，企业可以减少对设备购置、维护等成本的投入，进而降低整体制造成本。

（4）推动产业升级

共享制造资源促进了企业间的合作与交流，有助于形成更加紧密的产

业生态，推动整个行业的升级和发展。

（5）绿色协同的创新制造模式

企业可以更有效地利用清洁能源和材料，减少全社会废物排放和环境污染，推动绿色制造的发展。

4.3 质量管控

工业生产中的质量管控是极为重要的环节。质量管控可分为质量检测和质量管理两大工作。质量检测是指用多种手段获取来料、产品生产过程和产品终检出厂的质量数据和相关信息，涉及大量的智能传感技术，其中视觉感知最具代表性。质量管理则是基于质量检测获得的数据和多种来源的质量相关信息，根据工艺知识，对“人”、“机”、“料”、“法”、“环”诸类要素进行调整、管理，实现质量最优目标。此外，设备的状态监控和运维也可纳入质量管控工作范畴。质量管控作为智能制造最重要的环节之一，是人工智能技术应用的重要方向。

AI大模型出现后，其多模态海量信息处理能力、强大的泛化能力、数据分析能力为质量管控提供了强有力的手段。针对工业领域应用特点和工业大数据优化得到的工业大模型，在质量检测、设备状态监控与智能运维、质量管理各环节都实现若干成功案例。

4.3.1 质量检测

现代工业生产对质量检测在速度、精度、复杂工艺适应性和智能化等方面，提出了越来越高的要求，以机器视觉为代表的智能化检测技术已成为智能制造技术体系最重要的组成部分之一。工业大模型的出现，为智能化检测技术提供了有力的支持，因此成为智能检测开发商研究的重点，当前其成功应用主要包括：

（1）实现检测算法快速迁移：基于深度学习算法针对每一种新的产品若要实现高性能的检测，需要新产品的大量训练数据和较长的训练调整

周期。而工业大模型具备较强的泛化能力，便于实现快速迁移。初步实践表明，基于工业大模型开发的检测算法，可以只需极少的新样本和很短的训练周期就能在新产品甚至新行业，表现出不错的检测性能，极大地降低了开发周期和成本。

（2）改善质量检测性能：初步实践表明，基于工业大模型底座开发的垂直行业检测模型，在复杂背景检测、弱缺陷检测方面具有较好的性能优势，能够有效提升缺陷分割准确率和缺陷分类准确率，且能在一定程度上实现零样本的异常检测

（3）作为样本增广工具使用：许多工业质量检测场景下，缺陷样本的获取较为困难，基于工业大模型进行样本增广，有助于缩短开发周期，提升质量检测仪器性能。

例如，思谋科技开发的IndustryGPT工业大模型积累了超过300万张工业图像，又例如，凌云光开发的F.Brain工业质检视觉大模型(LusterLVM-2B基础底座)，如图4-2所示，以面向3C电子、锂电、显示面板、印刷包装、光伏等行业，积累了超过1000万张典型行业的原始缺陷样本，形成了工业缺陷生成模型、工业缺陷辅助标注模型、工业缺陷提示检测模型等多种模型。。此外，工业生产过程异常复杂，涉及多种变量和不确定性因素，传统的感知、处理、决策与控制解决方案蕴含了丰富的专家经验与工业知识，为工业大模型的知识学习提供了强有力的支撑。

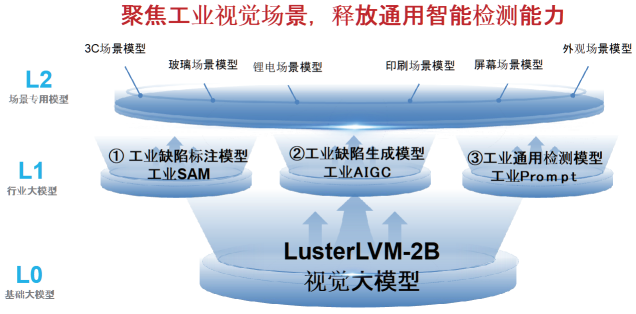

图4-2 F.Brain工业质检视觉大模型

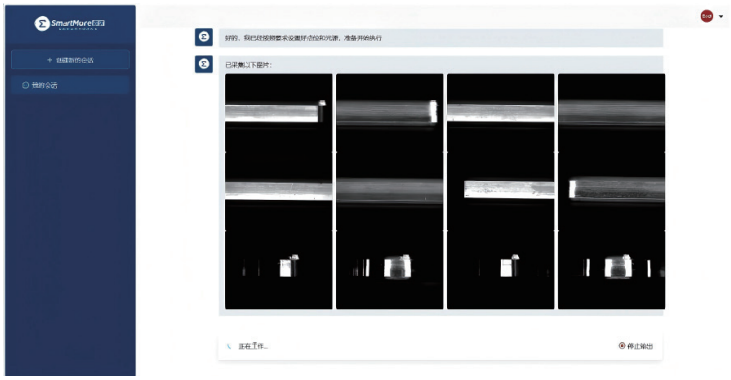

图4-3 思谋工业大模型服务平台

4.3.2 设备状态监控与智能运维

工业设备的运行状况和健康度直接影响工业生产的效率和质量。在智能制造体系中，对设备的运行参数、工况的监测可以获得大量的、多源的、多模态的实时数据，。基于这些数据和积累的历史数据，运用工业大模型可实现对设备状态的判断和预测，实现设备预测性维护，并在故障诊断和维修时提供辅助支撑。

4.3.3 质量大数据处理和质量管理

与质量相关的企业“人”、“机”、“料”、“法”、“环”数据，实时的质量检测结果，设备运行监测数据，以及历史数据和工艺知识共同构成了企业质量大数据和知识库，基于质量大数据，工业大模型可以有效改善企业质量管理水平。在尚未彻底解决大模型的可信赖性和可解释性的情况下，工业大模型可为质量管理提供辅助分析和决策。具体应用场景包括：质量回溯和根因分析：对出现的质量问题，基于大数据进行回溯，查找故障点和原因，并基于对工艺知识的分析，查找质量根因，便于快速定位问题源头，指导调整改进；质量预测预警：基于质量大数据，工业大模型可实现对生产质量的预测和预警，便于企业管理者提前安排对策并进行调整；质量工艺改进和良率爬坡：运用工业大模型对质量相关历史数据、实时数据和质量工艺知

识进行综合分析，指导调整生产工艺参数，改善产品质量。在新产品导入时，工业大模型可以快速寻找最优工艺路线，缩短良率爬坡周期。

4.4 物流配送

4.4.1 配送路径优化

工业大模型结合地理信息系统（GIS）、全球定位系统（GPS）和交通大数据，对物流车辆的行驶路线、运输资源分配以及配送计划进行智能优化。通过实时获取交通路况、车辆位置和状态、货物配送需求等信息，运用路径规划算法和智能调度技术，为物流车辆规划最优的行驶路线，提高车辆的满载率和运输效率，减少运输时间和成本。根据客户的地理位置分布、订单重量体积、配送时间要求等因素，合理安排车辆的配送路线和配送任务，实现快速、准确的货物投递。

4.4.2 装载优化

根据货物种类、体积、重量等多维度信息，建立装载模型并智能匹配合适的装载工具，提高集装箱或车辆的空间利用率，减少空载率；提高货物在运输过程中的稳定性从而降低货物损坏风险。

4.4.3 运输资源协同

基于货物特征、存储位置、运输需求、运输资源、路况信息、天气状况、成本效率等约束条件，应用工业大模型整合和分析海量物流数据，打破物流各环节以及产业链上下游间的数据壁垒，预测物流需求、统筹物流资源，开展供需匹配和协同运输，提高物流运输效率。

4.5 营销

4.5.1 客户画像与精准营销

整合企业内部的客户关系管理系统（CRM）数据、销售数据以及外

部的市场调研数据等，运用数据挖掘和机器学习技术，构建详细、精准的客户画像。基于客户画像，企业可以制定个性化的营销方案，如精准推送产品信息、定制化促销活动等，提高营销活动的针对性和推广效果，提升客户的购买转化率。例如，在高端装备制造行业，通过分析客户的行业类型、企业规模、采购历史等信息，大模型可以识别出潜在的大客户，并为其提供定制化的产品解决方案和专属的营销服务，增强客户对企业的信任和合作意愿。

4.5.2 销售预测与市场趋势分析

通过对宏观经济形势、行业发展趋势、竞争对手动态、产品生命周期以及企业自身销售数据等多源数据的分析，工业大模型应用时间序列分析、回归分析、机器学习等方法，对产品的销售趋势包括销量、销售额、市场份额、价格走势等关键指标进行预测。结合市场趋势分析，帮助企业提前洞察市场变化和潜在机会，制定相应的市场策略和销售计划。

4.6 售后服务

4.6.1 智能故障诊断与远程技术支持

通过与现场工程师的交互、多语言知识问答、语音播报和数字人交互等形式，提供实时的故障诊断建议，解决传统依赖人力巡检模式所带来的漏检、少检、人员安全等多方面问题。理解客户的需求和问题，并提供相应的解决方案或指导客户进行相关操作。可作为智能客服，能够回答客户关于产品使用、维护、保养等方面的问题，提供及时、准确的信息支持。此外，大模型可以与客户进行流畅的对话交流，提升客户使用体验。

4.6.2 数据分析与决策支持

对售后服务数据进行深度挖掘和分析，发现潜在的问题和趋势，进行

产品寿命智能预测。包括客户反馈、维修记录、设备性能参数等。通过分析，企业可以获取有价值的信息，为决策提供有力支持。例如，通过分析客户反馈数据，了解产品的质量和性能问题，进而对产品改进和优化。

4.6.3 知识库构建与更新

构建售后知识库，整合和存储大量的工业知识和经验。为企业员工提供便捷的知识获取渠道，帮助其更好地理解和应对售后服务中遇到的各类问题。不断输入新的知识和经验并对知识库进行优化，实现知识库的实时更新和完善。

4.7 供应链管理

工业大模型在可帮助企业在不确定的市场需求和供应中不断优化供应链管理，有效为库存管理、仓储配送、运输管理、供应商管理等业务环节提供洞察力和优化空间。应用工业大模型可深入把控供应链各环节各数据的相互影响规律，提高供应链管理效率，有效降低供应链成本，防范和降低供应链风险。

4.7.1 供应商智能评价与资源配置优化

对供应商的历史交易数据、产品质量数据、价格波动情况以及供应商的财务状况、生产能力等信息进行综合评估和分析，建立供应商智能选择模型。根据品类差异构建模型并深度分析供应商交期、质量、服务、可靠性等绩效数据，开展供应商绩效评价，通过差异化采购策略优化配置供应链资源。

4.7.2 需求预测与供应链协同

应用工业大模型可分析历史数据、市场趋势、社会经济、政策法规、技术发展等因素，对未来市场和供应需求进行预测，从而优化调整供应链

计划和策略。通过平台实现企业与供应商之间的信息共享和协同工作，如实时共享生产计划、库存水平、物流信息等，供应商也可以根据企业的需求及时调整生产和配送计划，提高供应链的协同效率和响应速度。

4.7.3 风险识别与延伸管控

应用工业大模型可分析识别潜在的风险因素，如技术风险、气象风险、市场风险、政治风险等，触发风险处置预案及时采取应对策略，同时监测并集成上下游的工艺过程、制造质量、使用过程、售后服务等数据，应用大模型识别、分析并优化潜在的质量问题，降低供应链风险，持续改善供应链各节点的质量，实现质量延伸管控。

4.7.4 库存和库位优化

应用工业大模型可监测并分析供应链全链条各环节实时库存及趋势数据，结合市场需求预测并优化库存配置，提高供应链的透明度、敏捷性和韧性。同时根据货物的体积、重量、形状等特征参数与存储设备、存储要求、业务需求等因素建立模型，对仓储库位进行优化，优化库位布局和存储位置提高仓储利用率，降低库存成本。

4.8 企业管理

4.8.1 人力资源管理

在人力资源管理场景中，智能问答系统的应用可提升HR部门的工作效率，同时改善员工的工作体验。通过利用大数据和自然语言处理技术，系统能够实时响应员工关于福利政策、培训资源、考核标准等企业办公制度方面的查询，从而使人力资源管理更加高效和透明。

4.8.2 企业文档管理

基于企业需求，快速生成各类管理文件，如报告、计划、政策文件

等，工业大模型根据输入的数据和信息，自动填充模板中的相应位置；根据文档的内容、类型、日期等属性，对管理文件进行智能分类和归档；帮助员工快速找到所需的管理文件，理解复杂的查询意图，提供精准的搜索结果；识别并保护敏感信息，如客户数据、财务数据等，通过加密和权限控制等技术手段，确保信息的安全性和隐私性。

4.8.3 法律文件生成

基于案件的基本信息、事实情况和诉求点，自动生成结构完整、内容准确的法律文书；模拟律师的对话方式，引导企业逐步细化问题，提供更具针对性的法律指导，协助公司法务部门对合同进行审核、修改和跟踪，识别出合同中的关键条款、风险点和潜在的法律问题，从而提供修改建议；协助公司法务部门处理知识产权的申请和维护工作，如商标、专利的申请和续展等。

4.8.4 公司党建

构建包含党的历史、党的政策、党的纪律等方面的内容的党建知识库，为公司员工提供便捷、准确的党建知识普及和教育服务；建立党员信息管理系统，实现党员信息的数字化、智能化管理；推动党建与业务的深度融合，找出党建与业务之间的关联点和切入点，实现党建与业务的相互促进和共同发展；通过对党员的学习情况、活动参与度、工作表现等方面的数据进行统计和分析，客观地评估党员的表现和贡献。

4.9 环保

4.9.1 环境监测

通过对海量的环境监测数据进行分析和处理，大模型能够快速准确地识别出环境数据中的异常情况，为环境管理提供有力支持。

（1）在环境监测系统中，深度学习模型可以对传感器采集到的数据进行实时分析，自动识别数据中的异常值和趋势变化，提高监测数据的可靠性。在空气质量监测中，能及时发现污染物浓度的异常变化，为环境管理部门提供预警信息。利用历史监测数据，时间序列分析模型能够预测环境参数的变化趋势，为环境管理部门制定应对措施提供参考。根据过去的监测数据，预测未来一段时间内污染物的浓度变化，以便提前采取治理措施。

（2）结合地理信息系统（GIS）和环境监测数据，空间分析模型可以对环境污染的空间分布进行分析，帮助环境管理部门确定重点污染区域和污染源。通过分析大气污染物的空间分布，确定污染源的位置和影响范围，为污染治理提供靶向目标。在环境监测设备管理方面，生成式预训练模型可以根据设备的运行数据和维护记录，自动生成设备的维护计划和故障诊断报告，提高设备的维护效率和可靠性。

4.9.2 污染治理

通过对污染治理工艺的模拟和优化，大模型能够找到最佳的治理方案，提高治理效果和资源利用率。通过传感器和一系列监测设备收集污染治理过程中的相关数据，包括污染物浓度、流量、温度、压力等，同时整合企业内部的生产数据、设备运行数据等。分析与模拟污染治理工艺与不同处理环节的效果和能耗，寻找最佳的处理工艺参数。识别污染治理过程中的约束条件，如排放标准、设备性能、工艺要求等，确保治理方案的可行性和合规性。

结合工艺分析、资源评估和约束分析结果，使用优化算法生成初步的污染治理方案。例如，在废气治理中，根据废气的成分和浓度，生成最佳的废气处理工艺和设备选型方案。将优化后的方案执行到污染治理过程中，并实时监控方案的执行情况，及时调整和优化治理策略。通过在线监测设备实时监测污染物的排放情况，根据监测结果调优治理工艺参数。

第五章 面向智能制造的工业大模型标准化现状与挑战

5.1 国内外标准组织

5.1.1 国际

当前在国际上开展大模型相关标准化工作的组织包括：

（1）ISO/IEC JTC1/SC42：主要承担JTC1中人工智能标准化项目，为其他委员会开发人工智能应用提供指导。国内对口组织TC28/SC42主要负责人工智能基础、技术、风险管理、可信赖、治理、产品及应用等人工智能领域的标准化研究工作。

（2）ISO/IEC JTC1/SC7：主要负责软件与 systems engineering 标准化工作。SC7涵盖软件生命周期、软件工程过程、软件质量和软件安全等方面。

（3）ISO/IEC JTC1/SC27：专注于信息安全、网络安全和个人隐私保护的标准化工作。

（4）ITU-T SG16：主要研究多媒体与数字技术，涵盖了音视频处理与压缩编码、多媒体系统及业务、数字文化、人工智能、大数据、区块链、车载多媒体系统及应用、数字健康系统及应用等标准化方向。

（5）ITU-T SG20：主要研究物联网、数字孪生和可持续智慧城市及社区，其目标是加速城市和农村地区的数字化转型。

（6）IEEE大模型标准工作组：于2023年11月成立，主要负责制定大模型技术规范、测评方法、安全可信、可靠决策等领域国际先进标准，为全球大模型产业技术创新和发展提供更好支撑。

（7）IEEE知识工程标准化委员会/知识图谱与大模型融合工作组：IEEE知识工程标准委员会于2023年6月25日正式成立，负责指导、管理和监督IEEE知识工程标准的立项、研制、评审、发布、宣贯及相关应用实践过程。目前，工作组围绕知识图谱与大模型融合、基于两者融合的知识服

务等标准化方向，也在开展相应的研制工作。

（8）CEN/TC 442:专注于人工智能和数据智能的标准化工作，涵盖大模型的开发、应用及其在不同行业中的集成。

（9）CENELEC TC 247:致力于工业自动化和制造技术的标准化，涵盖大模型在智能制造中的具体应用。

（10）ETSI TSI AI/ML:在人工智能和机器学习（AI/ML）领域开展了多项标准化活动，涵盖大模型在通信和网络管理中的应用。

5.1.2 国内

当前在国内开展大模型相关标准化工作的组织包括：

（1）全国信息技术标准化委员会人工智能分技术委员会（SAC/TC 28/SC 42）：主要负责人工智能基础、技术、风险管理、可信赖、治理、产品及应用等人工智能领域国家标准制修订工作，国际对口ISO/IEC JTC 1/SC 42。

（2）国家人工智能标准化总体组大模型专题组：承担大模型标准化制订工作，推动大模型技术和标准化的实践结合，促进人工智能产业健康发展。

（3）全国智能技术社会应用与评估基础标准化工作组（SAC/SWG35）：主要负责智能技术社会应用中的基础、通用、原则、测试方法、优化方法和效果评估等标准的制定。

（4）全国网络安全标准化技术委员会（SAC/TC 260）：是网络安全专业领域从事标准化工作的技术组织，对网络安全国家标准进行统一技术归口，统一组织 申报、送审和报批，具体范围包括网络安全技术、机制、服务、管理、评估等领域，同时开展人工智能安全相关标准的研制工作，国际对口为ISO/IEC JTC 1/SC 27。

（5）中国人工智能产业发展联盟：立足于搭建全球化的人工智能生

态合作平台，支撑政府决策，推进技术创新与应用落地，促进我国人工智能产业有序发展。

5.2 标准化进展

5.2.1 概述

当前，国内外已发布和在研的相关标准主要集中在模型评价、应用指南、通用大模型、深度学习、知识图谱等方面，暂无工业大模型的标准。就标准化方向而言，国际方面着重聚焦人工智能风险管控和伦理合规；国内方面则更加注重新技术落地应用。

5.2.2 国际

图 5-1 人工智能相关国际标准分布

当前已发布或者在研的国际标准如图所示，主要包括以下方面：

（1）在ISO/IEC JTC1方面，与大模型相关的标准包括：

a. ISO/IEC 42001提供可认证的人工智能管理体系框架，同时提出如何对人工智能管理体系的控制措施等内容。

b. ISO/IEC 38507:2022为人工智能数据的治理以及如何在人工智能系统整个生命周期内对其进行管理提供了结构化建议，还明确了人工智能的性质和机制，帮助管理机构理解使用人工智能的治理影响。

（2）在ITU方面，与大模型相关的标准包括：

a. ITU-T F.FDM-AC-BK“Assessment criteria for foundation models: Benchmark”（基础模型的评估标准：基准测试），提供了基础模型评估中基准评测的参考架构、技术要求和评估方法，包括模型能力、测试数据集、测试方法以及测试工具四个部分。

b. ITU-T F.TE-RAG“Requirements and evaluation methods for retrieval augmented generation of large scale pre-trained model”（大模型检索增强生成技术要求与评估方法），围绕RAG全生命周期过程中的技术能力和应用能力，从知识库构建能力、检索能力、生成能力、优化能力、应用成熟度、应用稳定性等维度进行展开。

c. ITU-T F.TE-AIA“Requirements and evaluation methods of artificial intelligence agents based on large scale pre-trained model”（基于大模型的智能体能力要求与评估方法），针对智能体产品和应用提出能力要求及评估方法，包括感知认知、规划、记忆、行动四个方面。

d. ITU-T F.TE-CG“Technical requirements and evaluation methods of AI based code generation in multimedia applications”（基于人工智能的代码生成技术要求和评估方法），围绕代码大模型相关的通用能力、专用场景能力和应用成熟度，主要从输入多样性、任务多样性、语言完备度、结果可接收性、结果准确度等维度，对代码大模型提出了全栈技术和管理要求。

（3）IEEE方面，与大模型相关的标准包括：

a.《人工智能基础模型能力定义及评测》：标准重点关注基础模型能力定义和测评，尤其是可评估的客观能力，旨在进一步指导人工智能基础模型的开发、测量和评估，以提高模型的透明度、公正性和有效性。

b.《大规模深度学习模型评估框架及流程》：标准旨在为大模型技术的可持续发展提供支撑，并推动大模型技术引领人工智能的快速创新。

c.《知识图谱与大模型融合框架》：标准旨在明确大模型（如大语言模型、多模态大模型）与知识图谱之间的通用融合框架，并提出大模型增强知识图谱、知识图谱增强大模型、知识图谱与大模型输出协同三种技术的对应技术框架。

（4）欧洲标准化委员会（CEN）方面，与大模型相关的标准包括：

a. EN ISO/IEC 22989:2023-人工智能概念框架:标准定义了人工智能的基本概念和术语，为大模型的开发和应用提供了统一的框架。

b. EN ISO/IEC 23053:2023–机器学习的概念和术语:该标准详细描述了机器学习的关键概念和术语，涵盖大模型的训练、验证和部署过程。

c. EN ISO 23894:2024–人工智能系统的伦理和社会影响:该标准聚焦于AI系统在社会中的伦理应用，并提出了大模型在智能制造中应用时的伦理指导原则。

（5）欧洲电工标准化委员会（CENELEC），与大模型相关的标准包括：

a. EN 61335-1:2023–工业自动化系统的安全要求：该标准规定了工业自动化系统中大模型应用的安全要求，确保系统在生产流程中的稳定性和安全性。

b.EN 62491:2024 – 大模型在生产流程优化中的应用规范：该标准详细规范了大模型在生产流程优化中的应用方法和技术要求，提升制造系统的智能化水平。

c. EN 62500:2024 – 工业物联网（IIoT）中大模型的集成标准：该标准旨在规范大模型在工业物联网环境中的集成和应用，促进数据互通和系统协同。

（6）欧洲电信标准协会（ETSI），与大模型相关的标准包括：

a. ETSI GS AI/ML for Network Optimization:2019–网络优化中的AI/ML应用指南：该指南为大模型在网络优化中的应用提供了技术规范，提升通信网络的智能化和自适应能力。

b. ETSI TS 103 645-1:2023–物联网设备的安全与隐私保护：该技术规范涵盖了物联网设备中大模型的安全与隐私保护要求，确保设备在处理敏感数据时的合规性和合法性。

c. ETSI TR 103 850:2024–5G/6G网络中的大模型应用规范：该报告探讨了大模型在下一代通信网络（如5G/6G）中的应用，并提出了相应的技术标准和实施建议。

5.2.3 国内

图 5-2 人工智能相关国家标准或地方标准分布

当前已发布或者在研的国家标准和地方标准如图所示，主要包括以下方面：

（1）国家标准

a.《人工智能 大模型 第1部分：通用要求》（国标计划号：20231736-T-469）：目的在于定义制备或使用大规模预训练模型的人工智能系统的

技术参考架构和相关方活动，并提出通用技术要求。

b.《人工智能 大模型 第2部分：评测指标与方法》（国标计划号：20231746-T-469）：目的在于定义预训练模型评测内容、指标设置和评测方法。

c.《人工智能 大模型 第3部分：服务能力成熟度评估》（国标计划号：20231741-T-469）：目的在于定义大规模预训练模型服务能力成熟度评估框架，

规定大规模预训练模型服务的能力要求、成熟度等级及评估方法。

d.《网络安全技术 人工智能生成合成内容标识方法》（国标计划号：20241842-Q-252）：描述了人工智能生成合成内容显式标识和隐式标识的方法，适用于规范生成合成服务提供者和内容传播服务提供者对人工智能生成合成内容开展的标识活动。

e.《网络安全技术 生成式人工智能数据标注安全规范》（国标计划号：20242097-T-469）：规定了生成式人工智能训练的数据标注基础安全要求、数据标注规则安全要求、标注人员要求、数据标注核验要求和标注安全测试方法。

f.《网络安全技术 生成式人工智能预训练和优化训练数据安全规范》（国标计划号：20242095-T-469）：规定生成式人工智能预训练和优化训练数据及其处理活动的安全要求，并提出对应的评价方法。

g.《人工智能 可信赖 第1部分:通则》（国标计划号：20240562-T-469）：该项标准为人工智能系统的可信赖性提供统一的框架和指导。

h.《人工智能 风险管理能力评估》（国标计划号：20231740-T-469）：旨在为人工智能（AI）系统的风险管理能力提供指导和评估方法的国家标准。

i.《人工智能 管理体系》（国标计划号：20221791-T-469）：该项标准为人工智能技术和应用的管理提供统一框架和规范。目标是帮助组织高

效管理人工智能项目和系统，确保其安全性、可靠性、合规性和可持续性。

j.《人工智能 边端设备模型部署工具链功能要求》（国标计划号：20242887-T-469）：该标准针对人工智能模型在边端设备（如物联网设备、智能手机、嵌入式设备）上部署所需工具链提出相应的功能规范要求。

k.《人工智能 多算法管理技术要求》（20232020-T-469）：该项标准规范多算法管理的技术框架与实现要求，确保算法选择、调度、组合等过程的高效性和安全性。

l.《人工智能 知识图谱 知识交换协议》（20230714-T-469）：知识图谱在跨系统、跨领域知识图谱的协作与共享面临诸多挑战，如数据结构差异、传输标准缺乏、互操作性差等问题。本标准定义知识图谱在不同系统之间的交换协议，确保数据的高效共享、可扩展性和互操作性

m.《网络安全技术 人工智能生成合成内容标识方法》（20241842-Q-252）：该标准目的是规范人工智能生成或合成内容的标识方法，为网络空间的内容管理提供技术支持。

n.《网络安全技术 人工智能生成合成内容标识方法》（20241842-Q-252）：该标准规范人工智能生成或合成内容的标识方法，为网络空间的内容管理提供技术支持。

o.《生成式人工智能技术应用社会影响 服务提供者合规管理指南》（20242884-T-469）：该标准旨在为生成式人工智能技术的服务提供者提供合规管理的基本框架和具体操作指南。

p.《生成式人工智能技术应用社会影响 评估指南》（20242891-T-469）：该指南旨在为生成式人工智能技术的开发、部署和应用，提供社会影响评估框架和方法，帮助各界更好地理解和承担这一技术的社会责任。

（2）地方标准

《工业人工智能平台 技术要求》（江苏省地方标准）：主要规定了工业人工智能平台的主要功能架构和各模块的技术要求。

（3）团体标准

《面向行业的大规模预训练模型技术和应用评估方法 第8部分:工业》：目的在于规范工业大模型在生产优化、分类识别、知识管理、生产管控、生产运营等场景的应用要求。明确工业大模型在感知、对话、分析、生成、决策等方面的技术能力要求。

《通用大模型工业应用服务能力要求》团体标准:主要关注当前通用大模型在工业领域应用对模型基本能力，以及研发、生产、管理等重点场景化应用能力的要求。

（4）其他

在大模型工业化应用成效评估方面，《中国AI大模型工业应用指数（2024年）》于2024年7月4日发布，指数体系由大模型基础应用能力与行业应用能力构成。大模型基础应用能力包含文生文、图生文等领域的准确性、稳定性能力；大模型的行业应用能力包含民爆、电力、石化、钢铁、医药等重点行业大模型在研发设计、生产制造、运维管理等环节具体场景的应用落地程度。

5.3 标准化挑战

5.3.1 工业场景复杂性相关的挑战

工业场景复杂多样，不同场景对模型的需求和性能要求各不相同，因此要求工业大模型在构建时需要充分考虑场景的多样性，具备足够的灵活性和可扩展性以适应不同的应用场景。这也导致不同场景的模型难以使用简单、统一标准进行规范和指导。同时，工业现场设备种类繁多，工业

协议难以相互兼容，国外引进设备接口不开放等问题导致设备数据采集困难。工业应用往往呈现碎片化的特点，即不同工业场景间差异较大且相互独立，这要求工业大模型能够快速适应和应对不同的碎片化场景需求，进而对模型在灵活性和适应性方面的标准化提出了更高要求。

（1）行业相关的挑战

行业多样性：智能制造涵盖机械制造、汽车生产、生物医药、新能源等各类细分行业。由于每个行业都有其独特的生产流程、数据标准和业务需求，工业大模型与通用大模型相比，需要适应不同行业的特定要求，增加了模型及其集成系统的设计和开发复杂性。

专业知识融合：工业领域各细分行业均拥有海量且独特的专业知识，如工艺流程、加工制造、材料转换等。工业大模型在应用到具体工艺流程时，需要能够准确理解和充分融合这些行业知识，但受限于制造业企业和行业内部知识资产积累和治理水平较低，知识来源分散度高，因此导致模型难以直接获得充足的知识源，且需面临知识深度融合等问题。

（2）多种设备相关的挑战

设备兼容性：生产环境中存在各种设备和传感器，涉及通信协议、数据格式和接口多样性。工业大模型在与设备进行集成应用过程中，需要能够兼容不同设备的接口和协议，进而实现数据的采集、传输和处理等操作，这对工业大模型对设备通信协议、数据格式和接口的兼容性提出了较高要求。

设备故障可用性：设备在运行过程中可能出现故障，影响生产效率和产品质量。工业大模型除了支持设备正常运行外，还需要具备对设备故障进行预测和维护的能力。这要求工业大模型与通用大模型相比能够学习必要的设备故障模式和维护方法，在典型工况下进行智能预警和维护建议，从而保障其在工业大模型的长期运行。

（3）工艺流程相关的挑战

工艺流程的复杂性：智能制造的工艺流程通常较为复杂，涉及多个环节和步骤。工业大模型需要能够理解和模拟整个工艺流程，包括原料准备、加工、检测、包装等各个环节。同时，模型还需要能够预测和优化工艺流程，以提高生产效率和产品质量。

工艺流程的变更和调整：随着市场需求的变化和技术的进步，工艺流程可能需要频繁变更和调整。工业大模型需要具备灵活性和可配置性，以适应工艺流程的变更和调整。这要求模型能够快速地学习和适应新的工艺流程，并更新相应的参数和算法。

（4）场景复杂性相关的挑战

场景的多样性：智能制造的应用场景多种多样，包括生产线自动化、物流配送、能源管理等。每个场景都有其独特的需求和挑战，如实时性、高可靠性、安全性等。工业大模型需要能够适应不同场景的需求，并提供相应的解决方案。

实时性和高可靠性要求：智能制造场景通常对实时性和高可靠性有严格要求。例如，在生产线自动化场景中，模型需要实时地监控设备的运行状态和生产数据，并快速地做出决策和控制指令。同时，模型还需要保证系统的稳定性和可靠性，以避免因系统故障导致的生产中断或产品质量问题。

5.3.2 数据相关的挑战

工业领域涵盖广泛，数据结构多样，包括41个工业大类、207个工业中类、666个工业小类，这将会导致数据质量参差不齐。高质量的数据是工业大模型训练和应用的基础，但现实中往往难以获得全面、准确、一致的数据集，这对于依赖大量数据训练的工业大模型来说是一个重大挑战。且工业场景下的产品整个全生命周期的数据（需求、设计、生产、营销等）之间缺乏相互关联，数据格式差异大，增加了数据收集的难度。且工

业数据往往包含大量噪声和异常值，需要长期进行复杂的数据清洗和预处理工作。这要求企业具备专业的数据处理团队和先进的数据处理工具，现阶段工业大模型训练样本数据质量差，缺乏抗干扰能力，对于一些对安全性和稳定性要求较高的场景而言，模型的可靠性难以得到保证。

（1）数据质量相关的挑战

a. 数据质量问题：在智能制造领域，数据质量直接影响到工业大模型的训练效果和预测准确性。然而，由于数据采集设备的精度限制、人为操作误差以及数据传输过程中的噪声干扰等因素，采集到的数据往往存在质量问题，如数据缺失、异常值、噪声等。这些问题会降低模型的训练效果，甚至导致模型无法正常工作。

b. 数据一致性和标准化问题：不同行业、不同设备产生的数据格式、单位、精度等可能存在差异，对数据的一致性和标准化造成挑战。为了构建通用的工业大模型，需要对数据进行统一的处理和标准化，以确保模型能够准确地处理不同来源的数据。

（2）数据多样性相关的挑战

a. 数据来源多样性：智能制造领域的数据来源广泛，包括传感器数据、设备运行状态数据、产品质量检测数据等。这些数据可能来自不同的设备、不同的生产线甚至不同的工厂。如何有效地整合这些多样化的数据，使其能够应用于工业大模型的训练，是一个重大的挑战。

b. 数据类型多样性：智能制造领域的数据类型也非常丰富，包括数值型数据、文本数据、图像数据等。这些不同类型的数据在描述设备状态、产品质量等方面具有不同的优势。因此，如何有效地处理这些多样化的数据类型，使其能够共同为工业大模型的训练提供支持，也是一个需要解决的问题。

（3）数据集相关的挑战

a. 数据集规模问题：工业大模型的训练需要大量的数据支持。然而，

在智能制造领域，由于数据获取和标注的成本较高，可用的数据集规模往往有限。这会影响模型的训练效果和泛化能力。因此，如何有效地扩大数据集规模，提高数据的质量和多样性，是工业大模型标准化过程中需要解决的重要问题。

b. 数据集质量和标注问题：高质量的数据集是训练出高性能工业大模型的关键。然而，在智能制造领域，由于数据标注的复杂性和专业性要求较高，数据集的标注质量往往参差不齐。这将会影响模型的训练效果和预测准确性。因此，如何确保数据集的高质量和标注的准确性，是工业大模型标准化过程中需要关注的重要问题。

5.3.3 技术演进相关挑战

随着工业技术的快速迭代，新的工艺流程、设备和技术不断涌现，工业大模型需要不断适应这些变化，进行模型的更新与升级，以确保其在新环境下的有效性和准确性，这一过程需要投入大量的时间和资源，且存在一定的技术难度。新技术的引入往往伴随着与现有系统的兼容性问题，工业大模型需要与不断更新的工业系统、设备和软件进行集成和协同工作，以确保数据的流通和模型的有效应用，不同系统之间的技术架构、数据格式和通信协议等差异可能导致兼容性问题，增加了模型应用的难度。同时随着工业大模型复杂度的增加和数据处理量的增长，其对算力的需求也在不断增加，提升算力需要投入大量的资金和资源，且存在一定的技术瓶颈，在工业领域技术快速演化的背景下，如何高效利用算力资源并降低算力成本成为了一个重要的问题。

（1）技术演化快速带来的挑战

a. 技术更新换代迅速：在智能制造领域，技术的更新换代非常迅速。新的算法、框架、硬件设备等不断涌现，使得工业大模型需要不断适应新技术的发展。这要求模型不仅要具备良好的兼容性和可迁移性，还需要能

够快速地进行更新和优化，以充分利用新技术的优势。

b. 技术标准化滞后：由于技术更新换代迅速，相关的技术标准和规范往往滞后于技术的发展。这导致工业大模型在标准化过程中缺乏统一的标准和参考，增加了模型设计和开发的难度。同时，也导致模型在不同行业、不同设备之间的兼容性和集成性面临挑战。

（2）技术复杂性带来的挑战

a. 模型复杂性增加：随着技术的不断发展，工业大模型的复杂性也在不断增加。模型需要处理的数据类型更加多样，包括数值型数据、文本数据、图像数据等。同时，模型还需要具备更强大的计算和推理能力，以应对更加复杂的工业场景和需求。这要求模型在设计和开发过程中需要充分考虑其复杂性和可扩展性

b. 技术集成难度加大：智能制造涉及多个技术领域，如人工智能、大数据、云计算等。将这些不同领域的技术进行有效集成是构建高性能工业大模型的关键。然而，由于不同技术之间的差异性和复杂性，技术集成难度加大。这要求模型在设计和开发过程中需要充分考虑不同技术之间的兼容性和协同性。

（3）技术可靠性和安全性挑战

a. 技术可靠性要求高：智能制造对技术的可靠性要求非常高。任何技术故障或缺陷都可能造成生产中断、产品质量问题等严重后果。因此，工业大模型需要具备高度的可靠性和稳定性，以确保其能够持续、稳定地为智能制造提供支持

b. 技术安全性挑战：随着智能制造的不断发展，网络安全和数据安全成为越来越重要的问题。工业大模型作为智能制造的核心组成部分之一，面临着严重的网络安全和数据安全挑战。如何确保模型的数据安全和避免黑客攻击和数据泄露是模型设计和开发过程中需要重点考虑的问题。

5.3.4 可解释性相关的挑战

多模态建模等原因带来的大模型复杂度增加，会导致模型的可解释性降低。工业大模型在捕捉数据模式时，可能会捕捉到关联关系而非因果关系。这些关联关系虽然对模型预测结果有影响，但并不足以解释产生结果的原因。如何区分因果关系与关联关系，并提取和验证可解释的因果关系是工业大模型面临的一个重要挑战。同时，数据不确定性会导致工业大模型在处理工业数据时，需要进行复杂的数据预处理和转换，同时这些处理和转换会涉及多个步骤和算法，每个步骤和算法都可能对最终结果产生影响，由于这些过程往往复杂且不透明，用户很难理解它们是如何影响模型输出的。

除此之外，工业大模型需要在不同的工业场景中有效工作，而这些场景往往具有高度的多样性和不确定性，如果模型的泛化能力不足，那么在不同场景下其解释性也会受到影响。例如，在某些特定场景下表现良好的模型可能在其他场景下表现不佳甚至失效。这都要求工业大模型在标准化过程中必须考虑可解释性这一关键指标。

5.3.5 可信赖相关的挑战

工业大模型在使用过程中容易出现幻觉现象，即输出不准确或误导性的信息。这在工业领域尤其危险，因为错误的预测或决策可能导致严重的生产事故或经济损失。例如，在化工领域，错误的化学反应路线或工艺参数可能造成化学品质量问题甚至安全事故。目前，通用大模型在知识体系测试中的表现往往低于工业领域对准确性的要求。因此，如何提升工业大模型的准确性和稳定性，防止幻觉产生，是一个亟待解决的难题。

2023年发布的《大模型可信赖研究报告》中指出：“大模型在快速发展的同时也带来了一系列潜在的风险和挑战，一方面，大模型所需的海量数据、复杂参数以及工程难度放大了人工智能固有的技术风险，如数据窃

取、泄露等安全问题，模型黑盒导致决策结果难预测和难解释问题，以及模型面对随机扰动和恶意攻击的鲁棒性问题。”对于工业领域而言，数据来源的可靠性和模型的决策偏见问题影响工业大模型的可信赖程度。工业大模型训练所用的数据包括文本、图像、视频等多模态数据类型，且数据溯源难以把控，可能来自设备操作手册、历史数据等采集数据，也可能来自采购、开源数据等多种渠道，部分数据缺乏验证，可靠性难以评估。

工业大模型的训练数据通常来源于现实世界，而现实世界中的数据往往存在不平衡和偏见。例如，某些群体在数据集中的代表性不足，或者某些特征被过度强调，都可能导致模型在学习过程中吸收并放大这些偏见。

a. 应用场景的复杂性：工业大模型需要应用于各种复杂的工业场景中，这些场景具有高度的多样性和不确定性，而且工业生产环境存在高温、高湿、强电磁等干扰因素。不同场景下的数据分布、任务需求等输入存在差异，使得模型难以适应所有场景并保持可信赖的性能表现。

b. 幻觉问题：工业大模型可能会产生“幻觉”，即生成与实际情况不符的预测或决策。幻觉问题往往是模型对数据的过度拟合或理解不足导致的，会严重影响模型的可靠性和可信赖性。

5.3.6 可控性相关的挑战

当前国内外大模型训练时多采用谷歌公司的Transformer模型等底层架构，国内尚缺乏自主创新的底层架构，这导致在大模型预训练方面只能“在别人的地基上盖房子”。且用于大模型训练的国产芯片与国外相比性能上也存在一定差距，这种差距部分源于国内在算力层面的不足，尤其是与英伟达的高端GPU芯片相比，国内在高性能计算芯片方面存在明显差距。

而且国际上部分顶尖的GPU芯片产品对国内企业断供，甚至提供的是“阉割版”的芯片，进一步加剧了国内在芯片技术上的差距。尽管国内部

分公司在AI芯片研制和生产方面取得了很大进步，但国产芯片尚未得到市场的广泛应用。此外，训练大模型需要的算力巨大，以往主要依赖数据中心、超算中心和云计算中心的算力基础设施，而随着大模型的兴起，算力需求急剧增加，尤其是对训练算力的需求增长幅度可达10倍甚至100倍。种种因素导致工业大模型也存在可控性相关的诸多挑战。

a. 技术不可控性：工业大模型在伦理道德、价值观上，都面临着从不可控变为可控的挑战。由于大模型的复杂性和非线性特性，其输出往往难以完全预测和控制，这导致在实际应用中可能产生不可预测的风险。

b. 算力与数据需求：工业大模型对算力和数据的需求极大，这增加了模型训练和维护的复杂性和成本。同时，数据的质量和来源也是影响模型可控性的重要因素。如果数据存在偏差或错误，将直接影响模型的输出和可控性。

c. 模型稳定性、鲁棒性与安全性：工业环境复杂多变，对模型的稳定性和鲁棒性提出了更高要求。如果模型无法在不同环境和条件下保持稳定的性能，将影响其可控性。此外，工业大模型本身也可能成为攻击的目标。如果模型被恶意攻击或篡改，将直接影响其可控性和可靠性。

5.3.7 隐私保护相关的挑战

工业数据往往包含企业的敏感信息和商业机密，如设备运行状态、生产工艺参数等，这些数据的安全性和隐私保护是工业大模型应用中的重要挑战，一旦数据泄露或被篡改，将会对企业造成重大损失。在工业场景中，不同企业和部门之间可能需要进行数据共享和传输，以支持大模型的训练和推理，在数据共享和传输过程中，如果缺乏有效的加密和访问控制机制，也容易导致隐私泄露。此外，一些训练算法本身也可能存在隐私泄露的风险，推理结果可能间接泄露用户的隐私信息，例如通过分析推理结果，攻击者可能推断出用户的某些敏感信息。

a. 数据收集与存储：工业大模型需要处理大量的数据，包括用户信息、生产数据、供应链数据等。这些数据在收集、存储和传输过程中都可能面临泄露的风险。如果数据保护措施不足，攻击者可能通过黑客行为窃取或篡改数据，导致用户隐私泄露。

b. 数据共享与流通：在工业生态系统中，数据需要在不同主体之间共享和流通，以实现资源的优化配置和协同创新。然而，同时也增加了数据泄露的风险。如果数据在流通过程中没有妥善的保护手段，可能会被未经授权的主体访问和使用，因此造成数据泄露。

5.3.8 测试评估相关的挑战

针对工业大模型的测试评估标准尚未完全建立，虽然学术界和工业界正在积极探索，但尚未形成广泛认可的标准体系。工业领域的复杂性和多样性使得制定统一的评估标准具有较大难度，不同行业、不同场景下的评估标准可能存在较大差异，难以形成统一的标准。目前，对于工业大模型的成熟度评级体系尚不完善，这在一定程度上限制了对其应用效果和潜在价值的全面认识，缺乏统一的成熟度评级标准，使得企业在选择和应用工业大模型时面临一定的不确定性。由于工业领域的复杂性，制定工业大模型的国家标准的过程难度较大，且从标准的制定到发布一般需要耗费几年的时间。

a. 评估方法的多样性：不同的评估方法可能产生不同的评估结果。如何选择合适的评估方法来准确评估工业大模型的性能成为了一个难题。

b. 数据隐私和安全：在评估过程中，需要确保数据隐私和安全。然而，由于工业数据的敏感性，如何在保护数据隐私的同时进行有效地评估成为了一个难题。

c. 数据质量参差不齐：工业数据质量参差不齐，数据结构和分布多样，这给模型的训练和评估带来了很大困难。低质量的数据可能导致模型

训练不充分或产生偏差，从而影响评估结果的准确性。

d. 可解释性难题：工业大模型通常具有复杂的结构和大量的参数，这使得模型的行为和预测结果往往难以被完全理解和解释。缺乏可解释性会导致评估者难以准确判断模型的决策依据和性能表现，从而影响评估结果的准确性和可靠性。

e. 性能评估的全面性：工业大模型需要在多个维度上进行评估，包括准确性、鲁棒性、稳定性、实时性等。然而，由于模型的复杂性，很难在所有维度上都达到最优表现。因此，评估者需要在不同维度之间进行权衡和取舍，以找到符合自身实际应用的模型。

5.3.9 可信赖相关的挑战

当工业大模型在工业生产中发挥到一定作用是，其预测和输出结果可能对企业生产安全和经济效益产生重要影响。工业场景复杂多变，随着新场景的出现，由于数据不足或偏差，大模型在新场景推理过程中容易准确度下降。为确保模型输出可信，企业需要完善质量监控与评估机制，对数据采集、标注和验证各环节进行严格把关。

在多模态及跨行业数据融合应用中，大模型可能捕捉到大量未知或无关特征，也会加剧预测结果的不确定性和偏差。由于大模型的“黑盒”特点，缺乏可解释性辅助信息。企业难以追踪模型推理过程，无法评估模型失效风险，将会导致决策者对模型的信任度降低。

保障工业大模型的可信赖性需多方协同，也是一项复杂多变的工作。在法规与标准层面，需制定清晰的责任与问责机制；在技术层面，需加强数据溯源、验证和安全审计；在应用层面，也需强化模型评估体系，并通过专家经验或业务规则对模型输出进行交叉验证。

5.3.10 幻觉抑制相关的挑战

工业场景中，生产流程往往环环相扣，存在严重的路径依赖。任何

环节出现偏差都可能引发连锁反应。大模型在面对多种来源的数据时，如果这些数据缺乏足够的处理，可能会导致模型无法理解它们的依赖和因果联系。因此，大模型可能产生幻觉，给出与实际情况相悖的结论。例如设备异常、工艺失衡以及外部环境的微小变化，都可能是产生这种情况的因素。模型输出的错误信息一旦被直接采纳，不仅会影响生产效率，还可能造成安全隐患。企业应建立定期的结果审查和反馈机制，既要针对数据和模型训练环节，也包括推理审核环节。以此来确保在大模型产生疑似幻觉时能够第一时间察觉并干预。

另外，在工业多源数据场景下，大模型可能识别到无关紧要但存在“表面现象”的信号，进而推理出根本不存在的因果关系。尤其当数据噪声和异常值较多时，更易引发虚假预测。针对这种风险，也需要对不同工业工艺和生产流程设计专项的鲁棒性测试，包括仿真环境测试、现场对比测试以及应急场景模拟，通过多重验证来识别和排除潜在的幻觉误判。

6.1 工业大模型标准体系框架

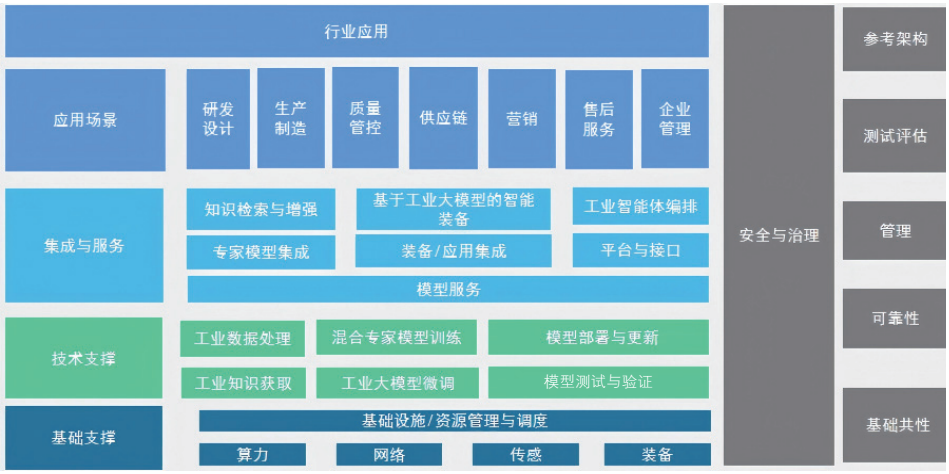

图 6-1 工业大模型标准体系框架

工业大模型标准体系结构包括基础支撑层、技术支撑层、集成与服务层、应用场景层等4个部分组成，如图6-1所示。

（1）基础设施层主要规范算力、传感、网络、装备、基础设施/资源管理与调度等技术要求，为工业大模型产业发展构建基础设施底座。

（2）技术支撑层相关标准主要规范工业数据处理、工业知识获取、混合专家模型（MoE）训练、工业大模型微调、工业大模型模型部署与更新、工业大模型验证等技术要求。

（3）集成与服务层相关标准主要规范知识检索与增强、基于工业大模型的智能装备、工业智能体编排、专家模型集成、装备/应用集成、平台与接口、模型服务等相关技术要求。

（4）应用场景层标准围绕工业大模型的生命周期，主要规范研发设计、生产制造、质量管控、供应链、营销、售后服务、企业管理等相关技

术要求。

6.2 工业大模型标准体系构成

6.2.1 基础设施标准

基础支撑层标准主要规范工业大模型主要规范算力、网络、传感、装备、基础设施/资源管理与调度等内容，主要包括：

（1）算力标准：主要规范面向工业大模型的算力性能、算力安全、算力评估、算力可靠性、算力稳定性等内容。

（2）网络标准：主要规范匹配工业大模型的通信协议、接口、性能指标（如带宽、延迟、吞吐量等）、网络架构的可扩展性等内容。

（3）传感标准：主要规范适配工业大模型的智能传感相关技术，包括测量范围与精度、稳定性与可靠性、响应时间等工业传感器的基本性能；规范通信接口与协议等数据处理与通讯标准内容；规范工业场景中工作环境与兼容性等环境与适应性标准；规范传感器测试与评估标准内容。

（4）装备标准：主要规范基于工业大模型的装备分类与定义、性能指标、设计标准、制造与加工标准、检测与验证、环境适应性等。

（5）资源管理与调度标准：资源管理标准主要规范资源库构建、资源分配、资源监测与优化等内容；资源调度标准主要规范调度策略与算法、负载均衡、多模型协同、容错与故障恢复等内容。

6.2.2 技术支撑层标准

（1）工业数据处理：规范数据采集、数据预处理、数据标注、数据分析、数据质量等；规范数据通信协议、数据交换协议、数据格式协议、数据采集协议、数据传输协议等。

（2）模型训练标准：规范模型架构标准；规范训练算法、超参数优化、训练配置、训练优化、并行与分布式训练等训练过程标准。混合专家

模型的设计原则、专家选择机制、专家模型融合策略等内容。

（3）模型验证标准：规范模型开发能力、数据处理能力、模型性能等技术能力验证标准；规范场景适应性、用户反馈等应用效果验证标准。

（4）工业知识获取：规范用于工业大模型的知识定义、知识获取、知识来源、知识表示、知识编码等内容。

（5）模型微调标准：规范微调数据要求、微调方法、性能评估指标、过拟合检测、可解释性分析和文档记录等内容。包括:明确微调数据的质量和来源要求,规定合适的微调策略,借鉴成熟的评估指标,检测并防范过拟合风险,分析模型的可解释性,并详细记录整个微调流程。

（6）模型部署与更新标准：规范模型部署所需的硬件、操作系统、运行依赖等环境要求，确保模型能够稳定运行；规范模型文件、配置信息等模型打包格式和方式，方便跨环境部署；规范明确模型服务的API定义、输入输出格式、错误处理等接口规范，确保与上下游系统的无缝对接；规范模型部署后需实时监控的性能指标，如响应时间、throughput、资源占用等性能监控指标；规范故障诊断、自动恢复、人工干预等异常处理机制、自动检测和处理方案等；规范模型触发条件、更新流程、版本管理、性能评估等。

6.2.3 集成与服务层标准

（1）知识检索与增强标准：规范知识表示与存储、知识增强方法、知识检索算法与接口、知识更新与维护等内容。

（2）基于工业大模型的智能装备标准：规范智能装备的功能与性能要求（智能感知与识别、决策与控制、协同与交互）、智能装备之间的通信协议和接口要求、场景适配要求、物理安全等内容。

（3）工业智能体编排标准：规范工业智能体的定义和分类、编排框架、感知和决策、智能体协同和交互、智能体功能和性能等内容。

（4）专家模型集成标准：规范专家模型的定义和分类、集成的原则和方法、性能评估指标和方法、接口协议、交互要求等内容。

（5）装备/应用集成标准：规范集成框架、功能模块标准、集成的基本原则和要求、数据格式等内容。

（6）平台与接口标准：规范工业大模型集成服务平台的总体架构、接口等内容。

（7）模型服务标准：规范服务能力、服务评估方法等内容。

6.2.4 应用场景标准

（1）研发设计标准：规范设计流程、数据来源、产品验证、专用数据库等内容。

（2）生产制造标准：规范工艺参数优化、设备监控、生产调度、生产流程优化等内容。

（3）质量管理标准：规范质量检测检验方法、质量管控流程、质量体系架构、质量、质量报告等内容。

（4）供应链标准：规范物流管理、库存控制、风险管理、信息共享、供应商管理、供应链安全等内容。

（5）营销管理标准：规范市场分析工具、客户画像、产品虚拟展示与交互、精准营销、客户数据生成、营销数据挖掘、营销效果评估、竞对研究等内容。

（6）售后服务标准：规范服务响应机制、售后流程处理、个性化服务框架、服务质量监控、物料快速供应、客户反馈收集、技术支持与培训等内容。

（7）企业管理标准：规范财务管理、人力资源管理、流程管理、战略管理、绩效管理、风险管理等内容。

6.3 工业大模型重点标准化方向

面向工业大模型的应用需求、技术实践现状及面临的挑战，当前工业大模型的重点标准化方向主要包括：

(1)《智能制造 工业大模型 通用要求》：提出面向智能制造的工业大模型在数据、模型、应用等维度的准确性、可解释性、可靠性、实时性等要求。

(2)《智能制造 工业大模型 测试评估方法》：提出工业大模型的性能评价模型、基准测试环境、评价流程及结果评价方法。

(3)《智能制造 工业大模型应用系统 参考架构》：提出工业大模型应用系统的统一参考架构，并明确其与传感器、装备、业务系统等集成框架。

(4)《智能制造 工业大模型 知识增强要求》：提出工业大模型与工业知识库间集成的框架，并给出各类知识增强类型的具体要求。

(5)《智能制造 工业大模型 工业智能体编排要求》：提出基于工业大模型的工业智能体结构，并提出工业智能体的编排流程及各环节要求。

(6)《智能制造 工业大模型 安全与治理要求》：提出工业大模型的安全防护、数据合规、隐私保护、应急响应等方面的具体要求。

(7)《智能制造 工业大模型 工业产品质量检测技术要求》：提出基于工业大模型的产品质量检测流程、性能要求、功能要求、测试方法等。

第七章 展望与建议

本章基于工业大模型的发展现状，给出了工业大模型的趋势展望，并提出了如图7-1所示的一系列针对性策略与建议。首先，聚焦于技术开发与应用，强调增强研发基础设施、推动产学研合作及创新技术监测评估体系的重要性。随后，深入探讨了标准制定与推广的必要性，包括加强标准制定与修订、推动标准应用及加强国际合作。同时，就政策支持与监管、人才教育与培养两大方面提出了建议，期望通过政策激励、监管框架构建及专业化教育体系等举措，为工业大模型的持续健康发展提供坚实保障。

图7-1 相关措施建议

7.1 趋势展望

在工业大模型的未来发展中，将有如图7-2所示的若干典型趋势。首先，细分领域的专用大模型将成为提升行业竞争力的重要驱动力。专用大模型能够深入理解特定行业的数据和工作流程，提供精准、定制化的解决方案；在此基础上，大模型与小模型的集成策略将进一步提高系统的响应速度和精度，推动智能化操作；与此同时，大模型与装备和应用的深度集成也将促进自动化生产和管理的发展，推动行业向数字化、智能化转型；开源与开放创新的趋势将为全球开发者和企业提供更多合作机会，推动技术突破；而通过知识增强的大模型建设，工业领域将能够更好地应对复杂数据和决策挑战，提供更智能、更高效的解决方案。以下将对这些趋势展开详细阐述。

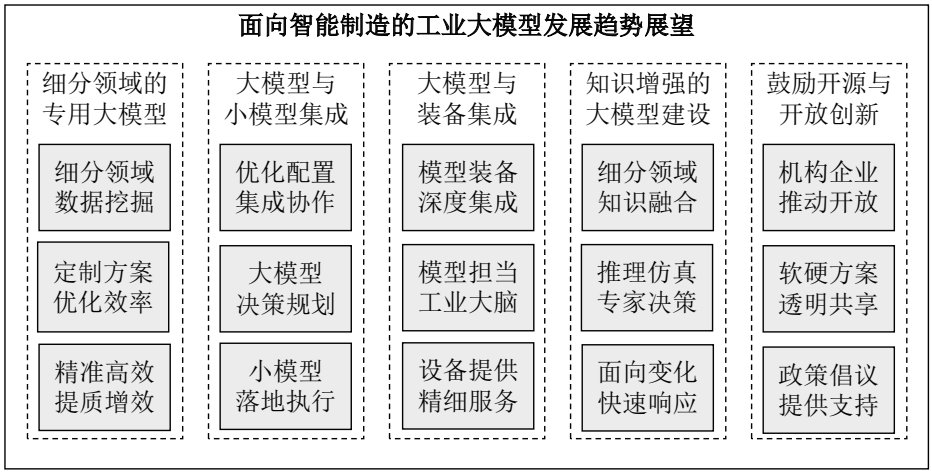

图7-2 发展趋势概览

7.1.1 细分领域的专用大模型

细分领域的专用大模型开发极具潜力，它能够针对特定工业领域的问题提供更精确和高效的解决方案，是提升行业竞争力的关键。这些模型在

深入学习特定细分工业场景的数据和工作流程后，能够提供定制化的解决方案，优化生产效率，提升服务质量。例如，在汽车制造行业，专用大模型可以优化车辆研发流程和产线工作节拍；在能源行业，专用大模型可以优化能源分配和预测电网负荷。未来，这些专用大模型将更加普遍地应用于各行各业，从而推动相关技术革新和经济增长。

7.1.2 大模型与小模型的集成

在实际应用中，大模型和小模型的集成提供了一种有效的解决方案。大模型凭借其强大的数据处理和分析能力，可以负责策略规划和长期决策；而小模型则可以在此基础上进行实时优化和调整，提高操作的精准度和效率。例如，在生产质检系统中，大模型可以进行质检环节规划和关键特征分析，小模型则可以提供实时数据流处理和基础特征提取。这种集成策略不仅优化了资源配置，还提高了系统的响应速度和可靠性。

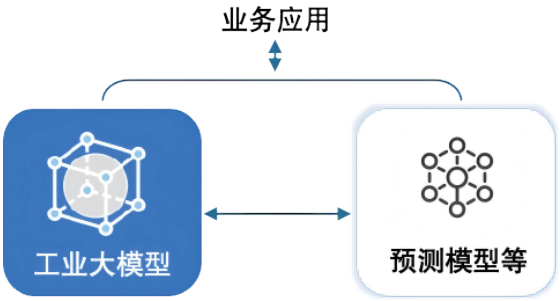

图7-3 大模型与小模型集成

7.1.3 大模型与装备/应用的集成开发

随着技术的进步，大模型将与更多智能设备和应用进行集成，形成高度自动化的生产和管理系统。这些集成系统通过高效的数据交换和实时处理能力，能够显著提高生产效率和产品质量。在制造业中，大模型可以集成到机器人系统中，优化生产线的自动化流程，并实现更精细的质量控

制。在服务行业，如零售和医疗，大模型可以与顾客服务机器人或智能诊断设备集成，提供更个性化和更高效的服务。随着云计算和物联网技术的成熟，这种模型与装备的深度集成将变得更加普遍，推动整个行业向智能化和数字化转型。

图7-4 大模型与装备集成

7.1.4 鼓励开源与开放创新

开源与开放创新是当前智能制造软件技术发展的主要驱动力之一。随着技术的快速发展，全球的开发者、研究机构和企业越来越倾向于共享其创新成果，推动科技领域的快速进步。开源不仅限于软件，也逐渐扩展到硬件和数据，使得复杂的技术解决方案更加透明和可访问。预计未来将有更多政策和倡议支持开源项目，尤其是在人工智能、大数据和物联网等领域。这种开放的创新环境不仅促进跨行业合作，而且为解决面向智能制造的生成式大模型技术突破提供新思路。

7.1.5 知识增强的大模型建设

为了处理日益复杂的数据和提高决策质量，将智能制造领域知识集成到大模型中逐渐成为一种趋势。将领域知识融入生成式大模型，使模型能够理解庞大的工业数据集，模拟专家的决策过程，在装备制造、矿山采掘、汽车质检等细分工业场景下提供专业的建议和解决方案，在不断变化的环境中做出快速准确的响应。

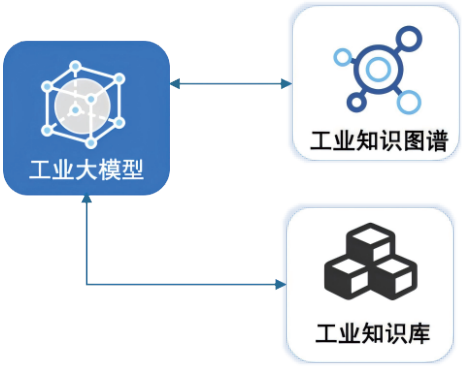

图7-5 工业知识增强的大模型

7.2 技术开发与应用建议

7.2.1 增强研发基础设施

建设强大的研发基础设施是推动技术开发与应用的关键。政府和私营部门应该投资建设高标准的研究实验室和创新中心，为科研人员提供先进的设备和资源。这些设施不仅能够加速技术研究的进程，还能吸引顶尖的科研人才。例如，可以设立国家级实验室，专注于智能制造、嵌入式软件、工业操作系统等关键领域的研究。同时，支撑建立企业研发部门与这些国家实验室的合作机制，促进科研成果的转化应用。此外，政府可以通过财政补贴和政策优惠支持中小企业建立自己的研发设施，增强整个行业的创新能力和竞争力。

7.2.2 推动产学研合作

为了加速智能制造大模型技术的开发与应用，需要加强产学研之间的合作。通过建立紧密的合作关系，可以使科研成果快速转化为实际应用，并且反过来，工业现场的需求也能反馈给学术界的人工智能研究机构。政府可以设立桥梁项目，如科技成果转化基金，专门用于扶持从实验室到市

场的转化过程。同时，鼓励高等院校和研究机构与企业共同开发课程，培养符合未来产业需求的人才。此外，建立行业标杆企业与高校的常态化合作模式，比如实习、讲座、联合研发项目等，进一步加深智能制造大模型理论与实践的相互结合。

7.2.3 创新技术监测和评估体系

在技术快速发展的今天，建立系统化的技术监测和评估体系显得尤为重要。该体系应持续跟踪国内外的智能制造技术发展动态，并进行定期的评估，以便及时调整企业和国家的研发方向。政府可以建立一个由专家组成的技术评估委员会，定期发布面向智能制造的工业大模型相关技术发展报告和政策建议。此外，该体系还应包括风险评估，特别是对于那些可能对社会、环境或经济产生重大影响的新趋势。通过这种方式，可以确保工业大模型技术的健康发展，减少潜在的负面影响。

7.3 标准制定与推广建议

7.3.1 加强标准制定与修订

智能制造作为制造业转型升级的关键方向，对于标准制定与修订工作提出了更高要求。首先，应当聚焦于智能制造核心技术和应用的需求，比如自动化设备、智能传感器、大数据分析、机器学习和人工智能技术等领域，制定相应的行业标准和技术规范。这一过程需充分考虑标准的前瞻性和灵活性，广泛征集行业内外专家的意见，整合多方技术视角和市场需求，确保标准制定的科学性和实用性。

其次，标准的修订也是保证技术更新和市场适应性的关键环节。应定期评估现有标准的适用性和有效性，根据智能制造技术的发展和行业应用的反馈，及时更新或修订标准。例如，随着人工智能技术在制造业中的深入应用，相关的数据安全和隐私保护标准需要不断地进行修订和完善。

7.3.2 推动标准应用与推广

在智能制造领域，标准的应用和推广是实现技术成果产业化和市场化的重要桥梁。首先，需要通过政策引导和市场激励，推动企业采纳新制定或修订的标准。政府可以通过制定优惠政策，比如税收减免、财政补贴等措施，鼓励企业实施标准化生产。同时，应通过建立标准认证体系，加强对智能制造标准实施的监督检查，确保企业在生产过程中严格按照标准操作。

其次，普及标准知识也是推动标准应用的关键。通过举办培训班、研讨会等形式，提高企业和行业从业者对智能制造标准的认知度和理解。这种培训不仅要涵盖标准的具体内容和技术要求，更应包括标准实施的最佳实践和案例分享，帮助企业更好地理解和掌握如何在实际操作中应用这些标准。

7.3.3 加强国际合作与交流

在智能制造的工业大模型标准化领域，国际合作与交流不仅是提升国际竞争力的关键手段，也是促进技术和经验共享的重要途径。随着全球制造业的迅速发展，国际标准在确保技术互通有无、推动全球市场一体化中起到不可或缺的作用。首先，应积极参与国际标准组织的活动。通过加入国际标准化组织（ISO）、国际电工委员会（IEC）以及其他相关的国际标准化机构，如国际工业互联网联盟等，加强国内智能制造在智能制造国际标准化舞台上的影响力。

其次，应积极举办和参与国际论坛、研讨会和展览会，加强与国际领先制造企业和研究机构的合作。这些活动不仅能展示国内智能制造发展成就，还能学习国际先进经验，建立广泛的国际合作关系。通过这种方式，可以有效提升国内智能制造标准的国际影响力，同时吸引国际资源和技术来华交流与合作，共同研发新技术，共享研究成果，提升国内和国际智能

制造的技术水平。

7.4 政策支持与监管建议

7.4.1 政策激励与资金支持

为了加速智能制造的发展，政府需要制定具有激励作用的政策和提供充足的资金支持。首先，政府可以设立专项基金，专门用于支持智能制造领域的研发和产业化。这类基金旨在减轻企业在初期技术研发和市场推广阶段的财务压力。其次，政府应当提供税收优惠，如减免研发投入的企业所得税、增值税以及关税等。这些政策不仅能够降低企业的运营成本，还能鼓励更多的投资流向智能制造领域。此外，政府可以通过政策引导银行和投资机构提供低利率的贷款和融资支持，特别是对中小企业。这类财政和金融政策将为智能制造技术的发展和应用创造一个有利的经济环境。

7.4.2 监管框架与合规指南

随着智能制造技术的快速发展，确立一个合适的监管框架至关重要，以保证技术发展的安全性和可持续性。政府需要制定明确的监管政策，指导企业在合法合规的框架内运行。例如，可以出台关于数据保护的法规，确保在智能制造过程中收集和使用的大量数据不会侵犯个人隐私或企业商业机密。此外，政府应当推动行业标准的制定和更新，如机器人操作的安全标准、人工智能系统的伦理标准等。这些标准将帮助企业评估并优化其智能制造系统，同时降低因技术问题导致的安全事故风险。监管框架的建立还应当鼓励公众参与和透明化，确保所有利益相关者能够对智能制造的发展方向有足够的了解和监督能力。

7.4.3 技术研发与创新合作

政府在推动智能制造的技术研发与创新合作中发挥着关键角色。为了

加快技术的进步，政府可以设立研发中心，聚焦于核心技术的突破，如物联网、大数据分析和人工智能等。同时，通过提供研发资金和建立研发联盟，可以激励高校、研究机构与企业之间的合作，共同解决智能制造过程中遇到的技术难题。此外，国际合作也是不可或缺的一环，政府可以通过签订双边或多边协议，引入国际先进的技术和理念，同时也将国内的创新成果推向国际市场。这种跨界合作不仅能够促进技术的快速发展，还能提升国内企业在全球智能制造发展中的竞争优势。

7.5 人才教育与培养建议

7.5.1 建立专业化教育体系

为了推动智能制造的工业大模型标准化，专业化的教育体系应发挥其重要作用。这个体系应涵盖从基础教育到高等教育乃至继续教育的各个层级。首先，高等院校需设置与智能制造相关的专业，如智能制造工程、机器人技术等，专注于培养和加深学生对智能制造系统的设计、优化和管理的深入理解。同时，职业技术学院也应开设相应课程，注重实用技能和操作技术的教学，以满足不同层次的人才需求。此外，校企产学研合作需要有针对性地强化，通过订单式培养、顶岗实习、校企研讨论坛、创业孵化等多种合作机制，实现理论教学与工业实践的相辅相成。

7.5.2 加强在职人员继续教育

随着智能制造技术的快速发展，现有工作人员需要不断更新其技术知识和技能。因此，制定面向在职人员的继续教育和培训计划至关重要。这包括定期的培训课程、研讨会和在线学习资源，特别是关于最新的智能制造技术、标准化进程及其应用的培训。企业应与专业培训机构合作，开发针对性的培训课程，帮员工掌握必要的技能。此外，鼓励员工参与国内外

的标准化活动，如标准委员会会议等，也是提高其专业水平的有效方式。

7.5.3 强化国际人才交流与合作

为了进一步提升智能制造的国际竞争力，强化国际人才交流与合作显得尤为重要。通过与国外高等教育机构和研发中心的合作，建立学生和教师的交换项目，不仅可以引入国际先进的教育资源和研究成果，还能提供机会让国内人才直接参与到国际智能制造的前沿技术研究中。此外，引进国外高水平的教授和行业专家来华授课、讲座，也是提升教育质量和技术水平的有效手段。同时，支持国内人才出国参加重要的国际会议、技术展览和标准化组织的活动，都是开阔视野、促进技术和标准化知识交流的重要途径。通过举办和参与这些活动，可以加强国际合作，促进全球资源的优化配置，提高我国智能制造领域在国际上的整体竞争力和影响力。

版权与免责声明

面向智能制造的工业人工智能与工业知识工程实践研究、标准研制及测试评估相关动态，敬请关注知识图谱与大模型公共服务网站、知识图谱标准化微信公众号、国家智能制造标准化总体组微信公众号或与联系人沟通。

知识图谱与大模型公共服务网站：
kgllm.cimsg.org.cn

国家智能制造标准化总体组微信公众号

知识图谱标准化微信公众号

中国电子技术标准化研究院

联系人：李瑞琪
联系方式：010-64102797
电子邮箱：lirq@cesi.cn