

综述与专论

DOI: 10.11949/0438-1157.20251408

通用AI对化工AGI的启示:数据-算力-工具-机理协同体系构建

顾紫怡, 王湘茹, 牛迎春, 王彧斐, 周天航

(中国石油大学(北京)重质油全国重点实验室, 北京 102249)

摘要: 人工智能正深度重塑工业生产范式, 化学工程领域中, 构建可自主理解、预测、设计与优化复杂化学过程的化工领域通用人工智能(chemical engineering artificial general intelligence, 化工AGI), 是推动行业颠覆性创新的核心目标。立足化工“三传一反”核心特性, 以数据-算力-工具-机理为脉络, 剖析化工AGI四大发展方向: 数据层面借鉴通用AI, 提出多尺度模拟驱动的自主造血式数据范式, 破解多尺度、高成本困境; 算力层面聚焦专用芯片与硬件架构, 论证化工专用计算机的必要性; 工具层面延续通用AI自主化逻辑, 探讨自动模拟、自动实验与自动生产控制对化工创新的助力; 机理层面强调物理化学机理与数据模型深度融合, 筑牢可信可解释根基。数据-算力-工具-机理协同形成自我强化生态, 聚焦行业需求与技术场景适配性, 为化工智能化转型提供系统性思考框架与发展蓝图。

关键词: 人工智能; 化工AGI; 三传一反; 算法; 模拟; 优化

中图分类号: TQ 028.8

文献标志码: A

文章编号: 0438-1157 (2026) 06-3239-19

Enlightenment of general AI to chemical engineering AGI: construction of synergy system integrating data, computing power, tools, and mechanisms

GU Ziyi, WANG Xiangru, NIU Yingchun, WANG Yufei, ZHOU Tianhang

(State Key Laboratory of Heavy Oil, China University of Petroleum, Beijing 102249, China)

Abstract: Artificial intelligence is profoundly reshaping the paradigm of industrial production. In the field of chemical engineering, the development of chemical engineering artificial general intelligence (chemical engineering AGI), which is capable of autonomously understanding, predicting, designing and optimizing complex chemical processes, stands as the core goal to drive disruptive innovation in the industry. Based on the core characteristic of “three transfers and one reaction” (mass transfer, heat transfer, momentum transfer, and chemical reaction) in chemical engineering, this paper analyzes the four key development directions of chemical engineering AGI with the framework of data, computing power, tools, and mechanisms. In terms of data, drawing on the experience of general AI, we propose a multi-scale simulation-driven self-sustaining data paradigm to address the dilemmas of multi-scale data acquisition and high acquisition costs. At the computing power level, it focuses on dedicated chips and hardware architecture, demonstrating the necessity of specialized computers for the chemical industry. At the tool level, it continues the autonomous development logic of general AI, exploring how automated simulation, automated

收稿日期: 2025-12-15 修回日期: 2026-06-01

通信作者: 周天航(1994—),男,博士,副教授, zhouth@cup.edu.cn

第一作者: 顾紫怡(2002—),女,硕士研究生, guziyi2975893694@outlook.com; 王湘茹(2003—),女,硕士研究生, 3158944862@qq.com

基金项目: 国家自然科学基金项目(22308376);上海市“通用人工智能大模型”基础研究专项(2025SHZDZX025D02);黄河三角洲学者项目(DYRC20230107);中国石油大学(北京)校青年拔尖项目(2462023BJRC015);学科前沿交叉探索专项(2462025XKQY002)

引用本文: 顾紫怡, 王湘茹, 牛迎春, 王彧斐, 周天航. 通用AI对化工AGI的启示:数据-算力-工具-机理协同体系构建[J]. 化工学报, 2026, 77(6): 3239–3257

Citation: GU Ziyi, WANG Xiangru, NIU Yingchun, WANG Yufei, ZHOU Tianhang. Enlightenment of general AI to chemical engineering AGI: construction of synergy system integrating data, computing power, tools, and mechanisms[J]. CIESC Journal, 2026, 77(6): 3239–3257

experimentation, and automated production control can contribute to innovation in the chemical industry. On the mechanism front, we emphasize the deep integration of physicochemical mechanisms with data models to consolidate the foundation of model credibility and interpretability. The four dimensions of data, computing power, tools, and mechanisms are mutually reinforcing and collectively form a self-evolving ecosystem. Focusing on the fit between industry demands and technical scenario adaptability, this work provides a systematic thinking framework and development blueprint for the intelligent transformation of the chemical engineering industry.

Keywords: artificial intelligence; chemical engineering AGI; three transfers and one reaction; algorithm; simulation; optimization

引 言

人工智能(artificial intelligence, AI)自诞生以来,一直试图模拟人类的认知能力。从图灵对计算本质的抽象思考,到神经网络对大脑神经元结构的仿生尝试,再到近年来大型语言模型(large language models, LLMs)在语言任务上表现出的涌现能力,每一次技术突破都在拓展AI的应用边界。通用AI在自然语言处理、图像识别以及化学合成、蛋白质结构预测(如AlphaFold)等任务上取得了显著进展。但“发现层面”的成功并不自动等同于“制造层面”的胜利。化工过程具有强非线性、非平衡、多尺度等特征,纯粹数据驱动模式并不完全适用,因此智能化路径必须结合化工机理与工业数据形成领域化方案^[1-2]。化学工程的核心在于对动量传递(momentum transfer)、热量传递(heat transfer)、质量传递(mass transfer)和化学反应(chemical reaction)进行精确调控,从而实现物质的高效转化与高值化利用。这一过程横跨从分子到宏观设备的多个尺度,涉及多物理场耦合,对预测精度、优化效果和控制鲁棒性都有很高的要求^[2]。传统化工研发与生产模式高度依赖工程师经验、物理化学理论及大量实验数据,因此具有知识密集、资本密集和迭代周期长的特点。面对复杂的分子设计、多变的工艺窗口以及日益严格的安全环保法规,传统模式的局限越来越明显。近年来,机器学习辅助分子设计、机制-数据混合建模以及机器学习势能模型等方法,为提升化工研发效率和过程优化能力提供了新的技术路径^[3-4]。

在这种背景下,通用AI的发展历程对化工领域的智能化转型确实有启发意义。但化工领域的特殊性决定了它不能简单照搬通用AI的技术路线:通用AI侧重数据拟合和模式识别,而化工领域通用人工智能(chemical engineering artificial general

intelligence, 化工AGI)必须与三传一反的物理化学规律深度适配,同时还要解决数据稀缺、计算复杂、可解释性要求高等特有难题^[5]。更关键的是,化工AGI的核心任务是解决从实验室发现到工业规模生产转化过程中的关键瓶颈问题。在这个过程中,非线性传递规律、复杂流体混合动力学以及严格的工程经济学指标构成了巨大的障碍。因此,化工AGI的构建不只是为了做一个更聪明的“自动实验员”,而是为了打造一个具备过程系统工程(PSE)全局思维、能够应对宏观工业扰动并实现全自主运行的“数字化工系统总工程师”。本文所称“通用AI”,指的是面向自然语言、图像、语音等通用场景开发的AI技术及应用,以区别于面向化工等垂直领域的专用AI。通用AI的演进经历了以符号逻辑为早期代表的知识驱动阶段,以及以人工神经网络为核心的连接主义(数据驱动)阶段,目前正逐步走向二者优势互补的神经符号系统(Neuro-symbolic AI)等混合智能架构^[6]。这一演进轨迹恰好映射了化工领域的核心挑战:既需要基于物理化学定律的严谨机理推理,也需要海量数据支撑的规律洞察。但连接主义模型的“黑箱”特性与符号逻辑系统的泛化局限,决定了化工AGI不能简单套用通用AI的路径^[7]。基于此,本文认为构建化工智能的核心逻辑,是将通用AI的技术思想与化工领域的专业特性深度融合,形成以数据为燃料、算力为引擎、工具为载体、机理为根基的四大支柱体系。这四个支柱之间存在深度的协同关系:数据的规模化积累需要算力支撑,算力的高效释放需要工具赋能,工具的落地应用必须以机理为约束,而机理的迭代完善又需要通过数据和算力来验证。这种环环相扣的协同关系,正是化工智能化的一条独特演进路径。

需要特别说明的是,本文所提化工AGI并不是指一个能够跨领域通用的AGI系统,而是借用AGI中“自主理解、推理与决策”的理念来描述化工领域

AI的终极发展目标。具体来说,化工AGI是指在化工这一垂直领域内,具备自主理解复杂化学过程、预测多尺度行为、设计最优工艺方案、实时优化生产运行能力的领域通用智能体。它的通用性体现在对化工全流程(从分子设计到工厂运营)的覆盖能力,而不是跨出化工边界的泛化能力。与通用AI致力于在开放域(open-domain)实现广泛的日常认知能力不同,化工AGI聚焦于在特定工业约束下实现跨尺度的系统级工程推理(system-level engineering reasoning)。同时,它不同于当前化工领域的专用AI(只解决单一环节问题),更像一个能独立应对化工复杂场景的“全能专家”,从研发到生产再到供应链管理,能够自主统筹并解决多维度复杂问题。

本文围绕化工行业的实际需求,从数据、算力、工具、机理四个支柱展开分析:数据部分重点讨论如何获取适配化工场景的高质量数据;算力部分聚焦专为化工计算设计的芯片与硬件架构;工具部分探讨如何让AI自主服务于化工研发与生产;机理部分强调如何让AI真正理解化工过程的本质规律。上述通用AI从符号主义、连接主义到混合智能的演进脉络,化工领域多尺度、数据稀缺等独特挑战,以及化工AGI以三传一反为核心的四大支柱协同运

辑,可通过图1来直观表示。

1 数据燃料:化工多尺度数据的创新获取与利用

数据是AI的“燃料”,这一点在通用AI的爆发式增长中已经得到充分印证:高质量、大规模、多样化的数据往往是模型能力涌现的前提条件。有观点认为,数据不是工业的语言,而是工业的矿石——这个说法与化工领域的数据困境颇为契合。化工数据的特殊性远超通用AI所处理的图像和文本数据:它横跨从分子尺度到宏观设备的多尺度范围,具有多源异构的特征^[8]。传统观点往往将挑战归咎于获取的高昂成本和实验数据的绝对稀缺,认为难以直接套用通用AI的海量数据采集模式。因此,化工AGI的数据创新,不仅要借鉴通用AI的数据范式思想,更要构建“矿石冶炼”能力:通过机理约束和专用工具,针对化工场景设计专属的数据获取与利用方案^[9-10]。

通用AI的突破往往以数据范式创新为先导,其核心逻辑是通过数据的规模化积累与多样化融合,构建支撑模型训练的高质量数据集。不同领域的代表性实践为化工AGI提供了直接的思想借鉴,各实践的核心特点及对化工AGI的具体启示见表1。

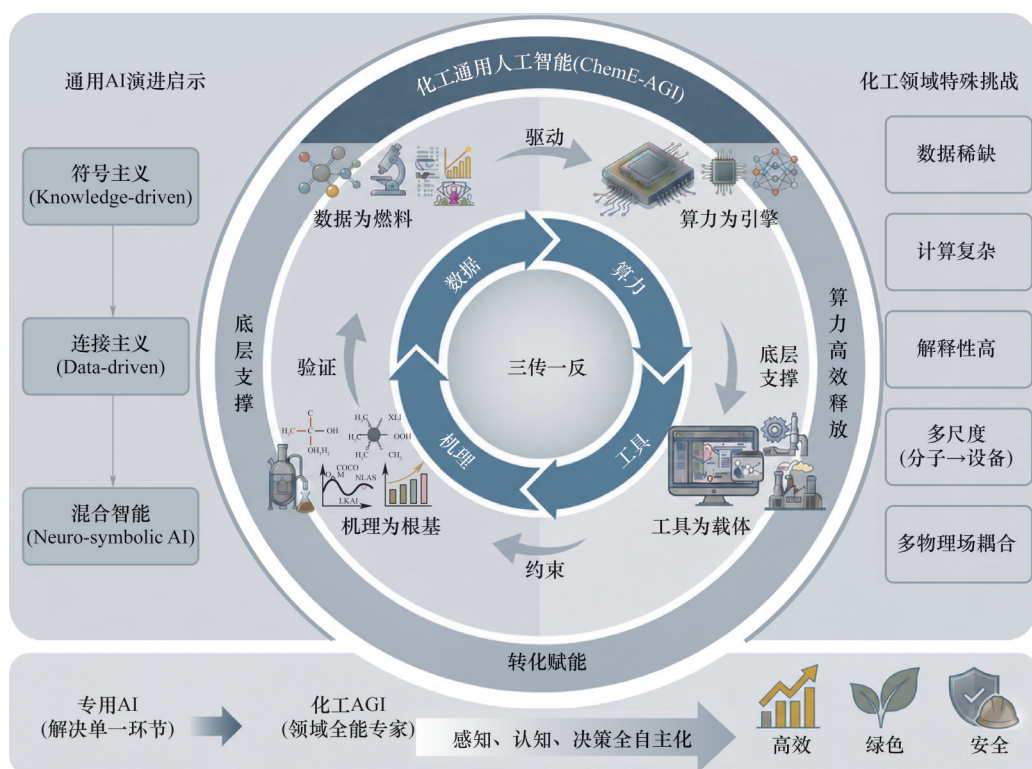


图1 化工AGI发展体系框架图

Fig.1 Development framework of chemical engineering AGI

表1 通用AI数据范式实践及化工AGI启示

Table 1 Practices in general AI data paradigms and their implications for chemical engineering AGI			
通用AI领域	代表性例子	核心特点	对化工AGI的启示
计算机视觉	ImageNet数据集	通过众包建立大规模、标准化标注数据集,定义了视觉任务的通用“数据语言”	建立跨尺度统一数据标准,夯实多源数据融合与模型训练基础
强化学习	AlphaGo训练体系	融合人类棋谱与千万级自我对弈数据,覆盖全场景博弈空间	依托少量高保真数据,驱动多尺度模拟生成合成数据,高效探索工艺参数空间
自然语言处理	GPT系列训练语料	整合跨领域、多模态、非结构化的原始语料,通过预训练学习通用表示	自动抽取实验报告、专利文献等非结构化信息,构建知识增强型数据集
语音识别	百度语音数据集	通过数据增强技术扩充数据多样性,提升模型鲁棒性与泛化能力	借化工机理模型扰动扩展实验数据,低成本丰富数据分布,提升模型泛化性

从这些案例中可以看出一条共性:通用AI的数据创新并不只是追求数据量的增长,而是通过模式创新解决数据的规模化、多样化和标准化问题。但化工领域的特殊性决定了它无法复制通用AI的众包数据范式,难点主要体现在三个方面:第一,化工数据标注不是简单的类别划分,而是需要建立温度、压力、浓度等工艺参数与产率、能耗、选择性等目标变量之间的结构化对应关系,这对专业知识的要求非常高;第二,化工实验具有强专业性、高成本和长周期特征——一次催化剂筛选实验可能耗时数周,大型设备运行数据的采集需要巨额资金投入,普通研究者或非专业群体无法参与数据生成;第三,化工数据横跨多尺度、多场景,数据可信度需要通过严格的实验设计和误差校准来保障,很难通过众包模式进行质量控制^[11]。计算机视觉领域ImageNet数据集的构建实践提供了一个重要参照:在深度学习发展的早期阶段,研究人员发现视觉模型的核心瓶颈是高质量标注数据的稀缺,从而通过众包方式构建了大规模层次化图像数据库。这一实践表明,在深度学习技术发展的特定历史阶段,数据范式的系统性突破往往能打破模型能力的瓶颈,其驱动意义与算法优化同等重要。然而,化工领域的特殊性决定了它无法简单复制众包模式。化工实验具有强专业性、高成本和长周期特征,一次大型设备运行数据的采集需要巨额资金投入。此外,化工数据标注需要建立工艺参数与目标变量之间极其复杂的物理对应关系,可信度依赖严格误差校准,非专业群体完全无法参与质量控制。然而,深入剖析时会发现,其根本制约并非仅局限于对领域专家经验、专业检测设备强依赖的表层限制,也并非来自真实工业场景中绝对的数据总量匮乏。事实上,连续稳定运行的化工生产装置,每日都会通过现场的DCS与SCADA系统,产生覆盖全生

产流程的海量时序运行数据。但这些在工业现场持续积累的原始数据,即便体量庞大,也与人工智能模型训练所必需的高质量、可直接复用的标注数据之间,存在着难以跨越的转化鸿沟。而这一鸿沟背后更深层的核心痛点,在于化工领域统一的数据表示与标准化框架的缺失。具体而言,当前业界尚未确立一套公认的标准化转换方法,无法将体系内复杂的流体力学参数、分子结构特征等专业信息,转化为化工垂类大模型能够高效识别与深度解析的物理语义格式^[12]。

因此,化工领域的核心数据矛盾,本质上是“未经机理清洗的海量工业数据”与“AI模型对高保真、结构化信息的旺盛需求”之间的严重失衡。解决这一结构性的数据困境,不能被动等待昂贵实验数据的缓慢积累,而应该主动创造并精炼数据。这需要构建一种基于化工领域技术特性的“自主造血”方案,即通过多尺度模拟技术,在虚拟空间中生成大量适配化工场景的理论数据^[1];同时,通过历史数据的高效机理化清洗机制,将DCS系统中的脏数据转化为结构化燃料。这个思路的核心,是利用化工模拟技术的成熟性,结合建模方法和计算技术的优化,构建规模化、标准化的化工数据集,从而在不依赖低门槛众包的前提下,为化工AGI模型提供充足且高质量的训练燃料,形成持续赋能的闭环生态^[13]。

粗粒化建模是多尺度模拟自主造血的基础。它的核心思想是在保留系统关键物理化学特性的前提下,通过简化模型表示、降低自由度来大幅减少计算量^[14]。具体来说,就是把多个原子或分子“打包”成一个虚拟的“超级粒子”,在不显著影响反应活性、热力学特性等核心性质的前提下简化计算,使原本难以完成的大规模模拟变得可行。在聚合物体系模拟中,Marrink等^[15]开发的Martini粗粒化力场将多个原子映射为单个珠子,显著提升了磷脂双

分子层和蛋白质-脂质体系模拟的时间和空间尺度,已被广泛用于生物膜和药物传递过程的研究。当前,粗粒化建模正在向智能化和自主化方向演进。Wu等^[16]提出可微分粗粒化方法(DiffCG),通过融合可微分模拟与多目标优化技术,实现了粗粒化力场参数的自主、高效优化。该方法可以同时精确匹配原子模拟中的多尺度结构特征与关键热力学性质,并成功构建了在宽温度范围内具有良好迁移性的聚苯乙烯熔体模型。DiffCG显著提升了多尺度数据生成的可靠性、效率与适用范围,代表了粗粒化建模从“参数化”迈向“自主优化”、从“单点拟合”拓展到“可迁移预测”的前沿进展,为化工AGI所需的高质量、跨尺度模拟数据提供了一种智能化的生成范式。

粗粒化模型可通过缩减系统自由度大幅提高计算效率,是多尺度模拟中常用的降维手段;但这类模型的简化处理均建立在特定假设基础上,不可避免地会丢失局部微观物理信息,最终导致模型预测能力具有强烈的“状态点依赖性”。这种仅通过尺度抹平实现的降维处理,无法真正跨越从微观量子、分子尺度到宏观流体、设备尺度之间的模拟鸿沟^[17]。而跨尺度模拟的核心壁垒,本质上源于结构描述的不一致性:化工领域的物理过程普遍存在多尺度耦合特征,但传统模拟软件与配套的计算硬件架构,大多仅针对单一尺度的模拟需求设计。针对这一跨尺度复杂问题的核心破解思路,来自介尺度科学(meso-science)的发展。中国科学院过程工程研究所李静海院士团队通过长期系统性研究证实,从分子、颗粒层面到反应器、生态系统层面,物质世界的不同层级均普遍存在特征性的介尺度结构,例如流化床内的颗粒聚团、多相反应体系中的气泡群等^[18]。这类介尺度结构呈现出的时空动态非均匀特征,其核心形成机制可归结为微观与宏观两套控制机制之间的“竞争中协调(compromise in competition, CIC)”原则。其中,以能量最小多尺度模型(energy minimization multi-scale model, EMMS)为代表的介尺度理论体系,为跨尺度模拟难题的解决提供了核心理论支撑与实现路径。在此基础上,当前前沿研究已实现介尺度物理先验知识与人工智能方法的深度融合,典型工作包括基于EMMS构建的ANN代理模型,以及介科学指导的深度学习(meso-science guided deep learning, MGDL)框架等^[19]。这类物理先验与数据驱动结合的融合方法,

既突破了传统粗粒化模型固有假设带来的局限性,显著提升了跨尺度模拟的预测精度与泛化能力,同时也为化工领域高性能计算突破冯·诺依曼瓶颈提供了核心的研发方向。

高通量计算是多尺度模拟自主造血的核心动力。它的本质是利用大规模并行计算技术,对成千上万种不同的化学结构、材料配方或工艺参数进行自动化、系统性的模拟^[20]。形象地说,它的运作模式类似工厂流水线,可以同时有成百上千种方案进行自动化模拟,快速筛选出最优选项,大幅降低人工试错成本。结合AI生成模型与自动化工作流管理软件(如AiiDA、FireWorks),可以构建强大的虚拟筛选与数据生成平台:Jain等^[21-22]通过高通量DFT计算构建了Materials Project数据库,包含超过15万种无机化合物的能量、能带结构、弹性性质等多维度数据,为材料设计与催化剂筛选提供了大规模标注数据集。在更微观的化学反应机制与界面过程研究中,高通量计算也借助机器学习势函数取得了突破。Cheng等^[23]针对电化学界面模拟的时空尺度瓶颈,发展了电化学机器学习势函数(ec-MLP)。该框架通过同时精确描述电解质的局域极化和金属电极的非局域电子响应,实现了对真实电化学界面的高效、高精度分子动力学模拟。这为研究三传一反中涉及电荷、物质与能量耦合传递的微观机理,提供了新的高通量数据生成工具,极大丰富了从分子到设备的多尺度数据来源。

这些数据横跨分子、介观、宏观等多个尺度,来源涵盖模拟计算、实验测量、文献挖掘等多种渠道。要把它们的价值最大化,关键在于整合与标准化——需要建立跨尺度的结构化关联,同时用统一的规范消除多源异质性,避免形成“数据孤岛”。在整合层面,可以采用非仿射晶格动力学等理论框架衔接不同尺度模拟的时间与空间间隙,或者通过贝叶斯数据同化技术融合多源观测数据与动力学模型。日本产业技术综合研究所(AIST)Sakaushi等^[24]在电催化剂研究中,就通过该技术整合电势、电流等多源数据,以仅300 h的实验数据准确预测了近1000 h的催化剂寿命,大幅缩短了降解评估周期;在跨尺度数据关联方面,机器学习方法已被用于实现微观分子模拟与宏观流场预测的耦合,为化工过程的多尺度建模提供了新途径^[25]。在标准化层面,可以借鉴统一数据模型(UDM)的设计思路,建立“反应物组成-过程参数-产物性能”全链条标准字段,

搭配JSON等结构化格式适配Aspen Plus、MFiX等主流软件。同时,参考美国国家能源技术实验室(NETL)在多尺度仿真框架中采用的分级管控机制,按精度将数据划分为量子力学级、实验级、文献级并标注可信度,为化工AGI筛选高质量训练数据提供依据^[26]。

多尺度模拟生成的虚拟数据,其核心定位并非替代实验测量数据,而是与实验数据形成互补协同的关系。虚拟数据可快速遍历实验手段难以覆盖的极端或宽范围参数空间,为人工智能模型的预训练提供充足的样本支撑;实验测量数据则承担了双重作用,既可以验证虚拟数据的物理可靠性,也能对模拟模型的核心参数与边界条件进行校准优化,二者结合最终形成“模拟生成数据-实验验证校准-模型迭代优化”的完整研究闭环^[27-29]。在多相催化反应的相关研究中,Tran等^[30]便采用了这一闭环思路:通过高通量密度泛函理论(DFT)计算生成了海量过渡金属表面的吸附能数据集,同时结合实验实测的催化剂活性数据完成贝叶斯校准,最终构建了可精准预测CO₂还原反应选择性的模型,相较于纯经验模型,该模型的预测误差降低幅度可达40%左右。而在聚合物加工领域,工艺优化的核心前提是建立能准确关联分子尺度本征行为与宏观制品最终性能的预测模型。针对这一行业核心需求,Mu等^[31]将粗粒化分子动力学模拟、原始路径分析与有限元方法深度耦合,搭建了一套覆盖多尺度层级的完整建模框架;以聚合物薄膜流延过程为具体研究对象,通过解析微观分子链缠结网络的动态演化,实现了对宏观熔体黏弹性的定量预测,同时还成功复现了实际加工过程中频发的典型缺陷,为利用计算模拟直接指导、优化聚合物加工全流程提供了可复用的完整范例^[31]。依托多尺度模拟形成的数据自主造血模式,研究人员可在较短周期内积累海量、高质量且覆盖全尺度层级的化工过程数据。这类数据集的核心优势在于,其不仅包含过程终点的性能指标,还完整覆盖了反应与传递过程中的全时空场数据与关键特征量,可形成“输入参数-演化过程-输出结果”强关联的结构化数据集。对于下一代化工AGI模型的开发而言,这类结构化多尺度数据是模型训练与泛化性验证的核心基础资源:它既能推动化工AGI从传统的数据驱动模式向深度数据智能模式演进,也将深刻重塑化工领域现有的数据生态,为构建可真正理解三传一反核心复杂过程的

化工AGI体系提供长期稳定的数据支撑^[32]。数据是化工人工智能模型迭代与能力升级的核心“燃料”,通用AI的数据范式带来的行业启示、化工领域当前面临的核心数据困境,以及上述数据自主造血方案与闭环生态的整体逻辑,均在图2中进行了系统呈现。

但海量化工数据的生成与处理,必然对计算能力提出很高的要求:数据作为燃料,其规模化积累与高效处理需要匹配的算力支撑。传统通用硬件很难应对化工模拟中复杂偏微分方程求解、大规模网格计算、多尺度耦合模拟等任务,无法满足自主造血数据范式对算力的需求。那么,如何构建专为化工多尺度模拟设计的算力体系,降低数据生成的计算成本、提升处理效率,让数据价值充分释放?这就需要化工专用算力引擎的技术突破。

2 算力引擎:专为化工计算设计的芯片与硬件架构

人工智能的每一次飞跃,都离不开算力引擎的驱动。算力是把数据转化为价值的核心动力,通用AI的发展已经证明:算法的突破必须与算力的升级同频共振,而专用化的算力架构往往能带来革命性的性能提升^[33]。对化工领域而言,算力的核心诉求不是通用计算能力的简单堆砌,而是针对三传一反计算特性的精准适配:化工过程的模拟与优化涉及大量非线性偏微分方程求解、稀疏矩阵运算和多尺度耦合计算,这些任务对硬件的并行计算能力、内存带宽、通信效率有独特的要求。通用CPU和GPU很难满足这些高效计算需求,因此,化工AGI的算力引擎必须是从芯片设计到硬件架构都专为化工计算定制的专用体系。在工业部署层面,算力支柱还需要构建“云-边协同”架构:在云端或企业级数据中心,利用强大的异构算力进行数字孪生模型的持续迭代与全局调度优化;在靠近生产设备的边缘侧,部署经过压缩和轻量化的推理模型,执行毫秒级的设备状态监测与闭环控制,以避免全量数据上传云端带来的网络延迟和网络安全风险。

通用AI的算力发展遵循需求牵引架构的逻辑。从通用硬件到专用芯片的演进,本质上是硬件对特定计算任务的精准适配。不同领域的专用算力实践为化工专用算力提供了直接借鉴,如表2所示,Anton超级计算机、NVIDIA GPU等通过架构适配特

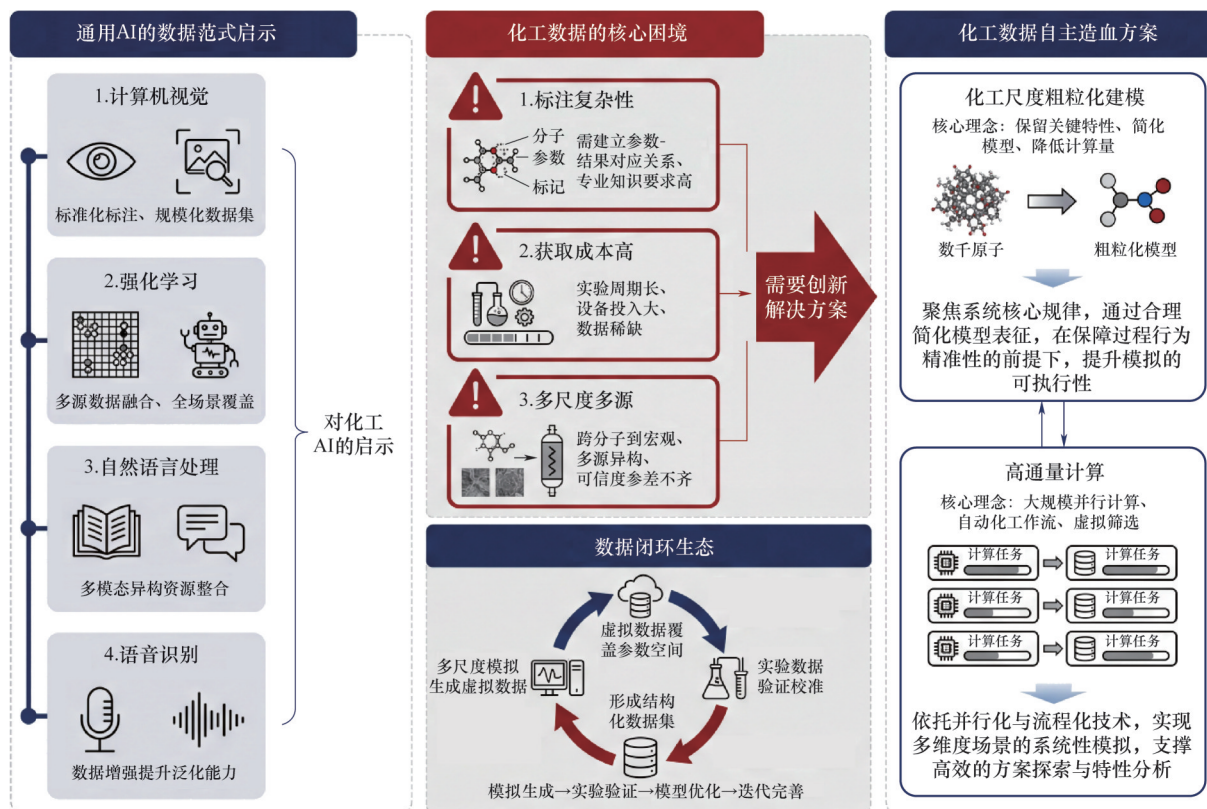


图2 化工AGI数据范式:通用AI启示、核心困境与自主解决方案

Fig.2 Data paradigm in chemical engineering AGI: insights from general AI, core challenges, and autonomous solutions

表2 通用AI专用算力架构及化工算力启示

Table 2 Specialized computing architectures in general AI and their implications for chemical engineering computing

通用AI领域	代表性算力架构	核心特点	对化工专用算力的启示
分子动力学模拟	Anton 超级计算机	专为 MD 模拟设计专用集成电路(ASIC)芯片,将核心计算烧录到硬件,性能比通用超算高两个数量级	证明了为特定计算模式定制硬件能带来颠覆性性能提升,化工计算可针对三传一反核心算法设计专用 ASIC
化工过程模拟	MIT 化工-CFD 专用加速芯片原型	针对多相流反应体系的 CFD 模拟,将压力-速度耦合求解器算法集成到硬件	直接印证了化工专用 ASIC 的可行性与巨大潜力,设计需紧扣三传一反的微分/代数方程组求解特征
深度学习训练	NVIDIA GPU (CUDA 生态)	针对并行计算优化,通过软硬件协同设计提升张量运算效率	化工专用算力需构建统一的软硬件协同生态,让工程师能通过高层描述(如方程或流程图)自动调用底层专用硬件
科学计算	美国劳伦斯伯克利 SSIC 原型	针对大气/流体动力学方程核心算子进行硬件加速,取得 10 倍以上能效提升	启示化工算力应对 Navier-Stokes 方程、反应网络求解等核心算子进行硬件级优化,而非通用计算单元堆砌
边缘 AI 应用	华为昇腾 AI 芯片	异构架构设计,兼顾通用计算与专用加速,适配低功耗场景	化工硬件可采用异构架构,适配模拟、优化、控制等多样化任务

定计算任务实现了性能跃升,印证了化工专用算力的设计方向。

上述案例揭示了一个规律:算力专用化的本质是硬件架构与领域计算特征的深度契合。化工过程的三传一反问题,其计算内核有鲜明的领域特性:计算流体力学涉及大规模稀疏线性方程组的迭代求解,对内存带宽和通信延迟极度敏感;反应动

力学模拟常常需要求解刚性常微分方程组,要求高精度数值积分;多尺度耦合计算则需要频繁的数据交换和尺度间映射。这些特征决定了通用 CPU/GPU 的架构并不是最优解。因此,化工专用算力的设计必须从化工计算的核心内核(如特定偏微分方程求解器、反应网络积分器)出发,进行从芯片指令集到硬件互连架构的全栈定制。算力专用化与协

同化是AI算力突破的核心路径,NVIDIA的技术实践是这一方向的典型代表。它通过软件框架与硬件架构的深度耦合来实现算力的高效释放——也就是用统一的编程接口衔接优化的并行计算硬件,构建可扩展的通用计算平台。这一技术路线推动GPU从传统图形处理器升级为通用并行计算核心,并依托CUDA编程环境落地:硬件层面持续优化并行计算架构,强化浮点数运算、缓存调度等核心能力;软件层面提供标准化编程接口,降低开发者利用GPU算力的门槛,最终使GPU成为深度学习革命的核心算力支撑^[34]。这一实践清晰地表明:当软件算法与硬件架构深度耦合、协同优化时,可以突破算力释放的瓶颈,实现数量级的性能跃升,这已成为算力转化为技术革新动力的核心原则。

这一思想对正面临算力瓶颈的化学工程领域有重要启示。化工的核心——三传一反问题,其内在的计算复杂性远超很多通用AI任务。这些问题通常由非线性偏微分方程组描述,求解过程涉及大规模网格计算、稀疏矩阵运算和多尺度耦合,具有极高的计算密度和并行度。传统的通用CPU虽然在处理通用逻辑和串行任务时表现出色,但在处理这类大规模、高并行度的计算密集型任务时,其冯·诺依曼瓶颈和标量计算架构导致效率和能效比都不够理想^[35]。以大规模CFD模拟为例,计算量随网格精度的提升呈指数级增长。在拥有数亿甚至数十亿网格的精细模拟中,即使是大型CPU集群,节点间的通信开销和巨大的功耗也会迅速成为难以逾越的障碍。化学工程领域是否也需要自己的高性能计算引擎,一个为三传一反物理过程模拟深度优化的专用算力架构? Anton超级计算机的成功实践为此指明了一条可行路径。Anton不是通用超算,而是专为加速分子动力学模拟从头设计的:它把MD算法中最耗时的核心计算(如原子间作用力计算)固化到定制ASIC中,再通过优化的3D环面通信网络减少数据交互延迟,最终实现了比通用超算高两个数量级的性能,把原本需要数年的蛋白质折叠模拟缩短到了数天^[36]。这种把算法与芯片、架构进行深度绑定的设计逻辑,为化工专用算力提供了关键借鉴——针对特定化工计算模式进行硬件层面的定制,可以带来突破性的效率提升。

由此延伸,化工AGI领域同样需要构建一台专为三传一反计算量身打造的专用化工计算机。它绝不是简单的硬件拼凑,而是从芯片设计到软件生

态的全链条协同适配,核心目标就是高效承载化工领域的计算任务——无论是CFD模拟、反应动力学求解,还是多尺度耦合计算,都能实现性能的突破性提升。要实现这个目标,其核心设计逻辑必须围绕化工计算特性来驱动硬件架构,从芯片、硬件架构到软件生态进行全维度定制化设计。从芯片层面看,需要设计化工专用ASIC芯片,把三传一反的核心算法(如压力-速度耦合求解器、刚性常微分方程组求解器)直接集成到芯片硬件中。在分子动力学模拟领域,日本理化学研究所(RIKEN)开发的MDGRAPE-4专用计算机系统,通过SoC架构将64个力计算流水线与65个通用核心集成于单一芯片,在生物大分子模拟中实现了比商用系统高一个数量级的模拟速度;在气候模拟和流体计算领域,领域专用加速器(DSA)通过将核心算子固化到硬件中,已展现出显著的能效优势^[37]。

从硬件架构层面看,专用化工计算机需要采用深度异构架构,集成化工专用ASIC芯片、通用CPU、GPU等多种计算核心,形成协同计算体系:CPU负责逻辑控制与任务调度,ASIC芯片处理核心化工计算任务,GPU辅助进行数据预处理与AI代理模型推理。这种架构的优势在于,可以根据化工计算的不同环节(如建模、求解、分析)灵活分配计算资源。欧盟“人脑计划”中开发的神经形态计算系统SpiNNaker2,集成了定制化的神经模拟ASIC与通用ARM核心,通过优化的片上网络实现多尺度脑模拟的高效能计算,其异构协同思想可以直接借鉴到化工多尺度耦合模拟中^[38]。在美国能源部支持下,橡树岭国家实验室打造的前沿超级计算机,采用CPU与GPU深度异构架构,并针对计算物理与化学应用优化了内存层次与通信库,在材料模拟与燃烧动力学计算中创下了性能纪录^[39]。美国劳伦斯利弗莫尔国家实验室等机构联合开发的分子动力学模拟系统,基于Cerebras WSE-2晶圆级芯片实现了80万个原子的高效模拟,速度较Frontier提升457倍,入围2024年戈登贝尔奖——这为化工专用算力芯片的架构设计提供了关键参考^[40];苏黎世大学团队则通过27块FPGA开发板构建低成本集群,针对分子动力学模拟优化互联拓扑与运算逻辑,其能效与模拟速度均超过了传统GPU集群,证明了FPGA异构架构在化工计算中的高性价比应用潜力^[41]。同时,化工领域的智能化转型需要以统一的软件架构与高性能计算平台作为底层支撑,以实现数据、模型与

算法的系统性融合^[42]。在此基础上,可以进一步借鉴科学计算领域成熟的软硬件协同思路,通过高层抽象与自动代码生成技术构建面向化工计算的工具链:让化工工程师不用关心底层硬件细节,只需要用高级语言(如Python)描述化工问题(如反应器内的流动与反应),软件就能自动把计算任务映射到最优的硬件单元上。例如,开源多物理场模拟框架FEniCS提供了高层抽象来描述偏微分方程,并能自动生成适配不同硬件(CPU/GPU)的高效代码,显著简化了化工传递过程模拟的开发流程^[43]。美国国家能源技术实验室开发的“多尺度仿真框架”MFIX,集成了流体力学、颗粒动力学与化学反应模型,并通过与高性能计算库(如PETSc)的深度集成,实现了在异构计算集群上的高效并行计算,为流化床反应器等典型化工设备的模拟提供了完整工具链^[44]。在自动实验平台领域,美国阿贡国家实验室基于机器人与AI构建的自主实验系统,5个月内完成了6000余次有机液流电池材料实验,效率比传统方法提升了10~16倍,其“AI驱动-实验验证-模型迭代”的闭环逻辑为化工AGI自主发现工具提供了国际范式^[45];北卡罗来纳州立大学开发的动态流控自主实

验室,通过实时监测反应过程实现了数据采集效率10倍的提升,在清洁能源材料与可持续化学品开发中展现出很大的应用价值^[46]。上述专用化工计算机的深度异构硬件架构设计、多计算核心协同逻辑及软硬件适配的软件生态构建思路,可以通过图3直观地呈现其完整协同体系。

专用化工计算机的构建,有望显著缓解化工计算的算力瓶颈:它让整座化工厂的高保真实时模拟成为可能,使工艺优化的计算周期从数月缩短到数天,为化工AGI提供强大的算力支撑。但算力的爆发式增长如果缺乏场景化工具来承接,其工程应用价值就难以充分释放。算力作为引擎,需要通过适配化工场景的智能工具,才能转化为研发与生产的实际价值。那么,如何让强大的化工专用算力自动服务于核心任务?这就需要化工专用认知工具的革新。

3 认知工具:让AI自主服务于化工研发与生产

认知工具是算力转化为化工价值的核心载体。在AI技术不断迭代的背景下,基础编程实现、工具

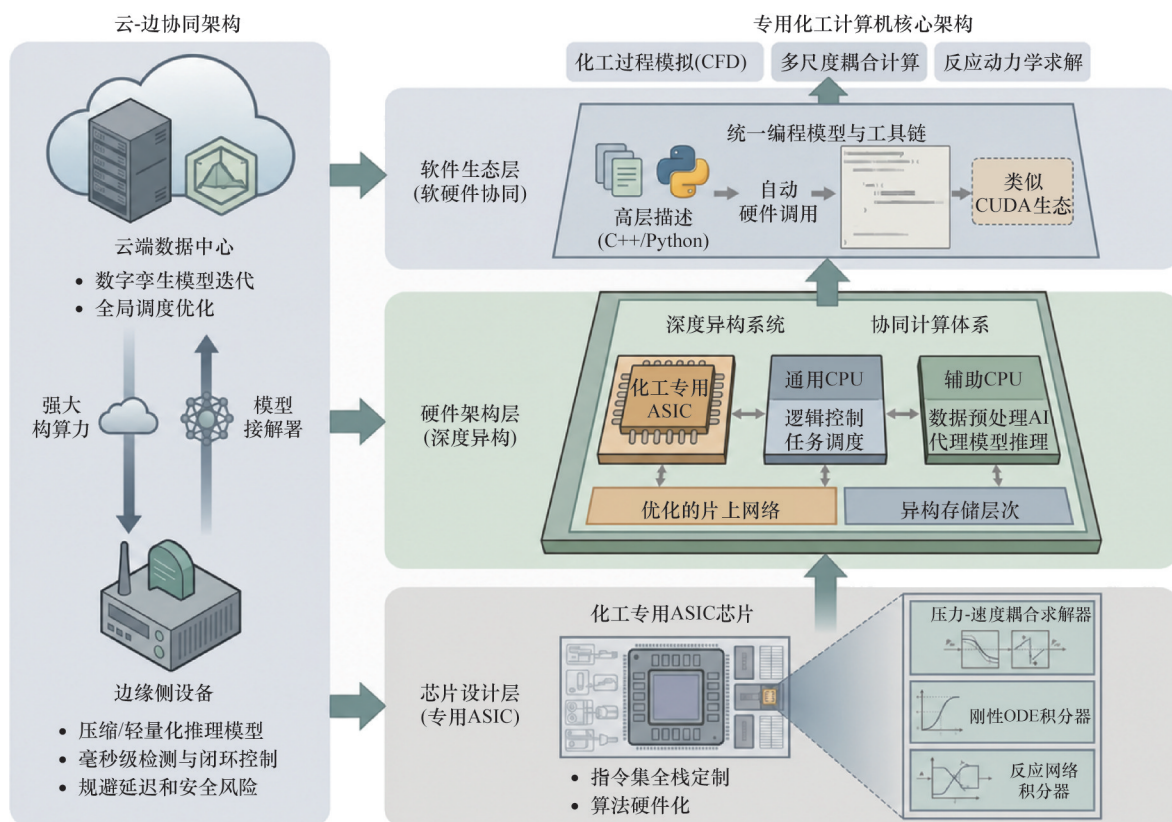


图3 化工AGI专用算力协同设计体系

Fig.3 Co-design architecture of chemical engineering AGI-specific computing systems

框架调试等技能的重要性在逐渐下降,而精准刻画场景特征、把抽象需求转化为明确指令的能力正成为核心竞争力。通用AI认知工具的演进始终围绕着自动化和自主性这条主线——从被动响应人类指令向主动探索客观规律升级,其自主决策与执行能力直接决定了AI技术在实际场景中的应用价值^[47]。对化工领域而言,认知工具的核心诉求不是复刻通用AI的全场景自主决策能力,而是要深度适配化工研发与生产的复杂场景特性,实现从人工操作驱动的计算工具到自动化驱动的智能辅助系统的转型,从而帮助工程师减少重复性劳动,把精力集中在工艺创新、流程优化等核心研发环节^[48-49]。

通用AI的工具发展,本质上是让机器适应人的需求——从被动执行到主动探索,其核心逻辑是工具对特定场景的深度适配。不同领域的工具实践为化工认知工具提供了思想借鉴,如表3所示,AutoML、机器人化学家等工具通过自动化、闭环迭代来提升效率,为化工AGI的自主模拟与实验工具开发提供了参考。

这些案例的共同启示是:有效的认知工具必须具备场景理解能力和自主执行能力——能够理解特定领域的需求,自动规划任务路径,独立完成核心工作。对化工领域而言,认知工具的场景理解能力意味着能够解读化工专业术语、工艺需求与约束条件(如最大化甲醇产率、控制反应温度不超过400 K);自主执行能力意味着能够自动完成模拟建模、实验设计、参数优化等重复性工作,形成“需求-执行-结果-反馈”的闭环。化工认知工具的核心目标,是让工程师从繁琐的建模、计算、实验操作中解放出来,聚焦于工艺创新、机理分析等更高价值的工作^[50]。LeCun^[51]曾提出一种追求自动化与自主性的工程哲学:不想通过烦琐的代码去硬编码智能,而是让机器自己学会智能。这背后是对工具自主

性的深刻期望:真正强大的AI工具不应该需要人类做大量底层操作,而应该能够自主理解目标、探索路径、解决问题。这一哲学恰恰映照出传统化工工具的局限性^[52]。传统化工工具(如Aspen Plus、COMSOL、实验室实验设备)本质上是被动计算工具,高度依赖工程师的手动操作:做CFD模拟时,工程师需要手动完成几何建模、网格划分、边界条件设置、求解器选择等一系列烦琐步骤,任何一个环节出错都可能导致计算失败;做工艺优化时,工程师需要基于经验设计实验方案,手动调整参数,记录结果,效率很低;做数据处理时,需要手动整合实验数据与模拟结果,进行误差分析与模型校准。这些工具的核心问题在于“人适应机器”,工程师需要花大量时间学习工具的操作逻辑,而不是机器去理解工程师的专业需求。

面对三传一反的复杂系统,传统工具的局限越来越突出:多变量耦合的工艺优化中,人工设计的实验方案很难覆盖全部参数空间;多尺度模拟中,手动调整不同尺度模型的接口参数非常耗时;生产过程的实时优化中,人工操作无法快速响应工况变化。化工领域需要的是LeCun所追求的自主工具——能够理解化工专业需求,自动完成建模、计算、实验、优化等工作,成为工程师的智能助手,而不是操作负担。这种自动化驱动的理念在化工领域的落地,体现为自动模拟与自动实验两大核心工具。这并非通用AI工具的简单适配,而是基于化工场景的专用创新,其核心是让AI自主服务于化工研发与生产的全流程,形成“模拟-实验-优化”的闭环。

自动模拟工具的核心是理解化工需求,自动完成建模与计算。传统的模拟工作流程,要求工程师具备深厚的计算流体力学(CFD)、计算化学或过程系统工程知识,并熟练掌握Aspen Plus、COMSOL、

表3 通用AI认知工具实践及化工工具启示

Table 3 Practices in general AI cognitive tools and their implications for chemical engineering tools

通用AI领域	代表性工具/技术	核心特点	对化工AGI的启示
自动化建模	AutoML(自动机器学习)	自动完成数据预处理、模型选择、超参数优化,降低建模门槛	聚焦化工机理约束,构建“模拟-实验-优化”闭环,自主适配模型并探索参数空间
自主实验	机器人化学家	闭环控制实现实验设计、执行、分析自动化,加速科学发现	推动多相流、强耦合体系的高通量自主实验,结合AI决策实现方案智能迭代
代码生成	GitHub Copilot	理解自然语言需求,自动生成代码,解放重复性劳动	实现化工专业场景的代码自动生成,覆盖模拟脚本、实验流程、数据处理等需求
强化学习应用	AlphaGo	自主与环境交互学习最优策略,无需人工预设规则	支持化工过程动态优化,自主应对反应条件波动、多目标约束等复杂工况

Fluent 等专业软件。工程师需要手动完成问题定义、几何建模、网格划分、边界条件设置等一系列烦琐步骤。与那些只能处理单一建模环节(如只负责网格划分或求解器选择)的化工专用AI不同,面向化工AGI的自动模拟工具深度融合Agent Skills技术,可以支撑化工AGI的全流程自主决策——工程师只需要用自然语言描述需求(如“设计一台固定床反应器,在400 K、5 MPa条件下催化CO加氢合成甲醇,最大化产率”),工具就会调用预存的化工模拟专用Skill包,通过经过实验验证的脚本(如Gaussian/LAMMPS输入文件生成脚本)与标准化规则自动完成建模与计算。这种做法既解决了传统大模型生成文件的“幻觉”问题(按原子数自动匹配计算方法/基组、参考官方手册校验语法、对照官方示例校准参数),又能通过机器学习优化网格密度等建模参数,在保证脚本物理合理、语法正确的前提下提高计算效率,而且具备模拟结果与实验数据出现偏差时的自学习修正能力,不需要人工介入中间环节。在这个过程中,AI的作用是把化工专业知识(如不同反应器的建模规范、反应动力学模型的选择标准)转化为自动化规则,同时通过机器学习优化建模参数(如网格密度、迭代步数),在保证精度的前提下提高计算效率。开源平台Simvue集成了机器学习与物理模拟,允许用户用高级语言描述化工过程,自动生成对应的CFD或过程模拟案例并优化计算参数,在催化反应器设计中显著降低了建模门槛。商业软件ANSYS Fluent通过嵌入AI辅助网格生成与求解器设置功能,可以根据用户输入的几何与物理条件自动推荐最优离散方案与收敛控制参数,提高了模拟效率与可靠性。此外,自动模拟工具具备自学习与自修正能力:当模拟结果与实验数据存在偏差时,工具会自动分析偏差来源(如动力学模型参数不准确、边界条件设置不合理),调整模型参数或建模方案,实现模拟精度的自主提升,这也是化工AGI全流程自主化能力的核心体现。

自动实验工具的核心是自主探索工艺参数空间,优化实验方案。化工实验的核心目标是在多变量耦合的参数空间中,找到最优操作条件(如温度、压力、进料配比)。传统的实验设计方法(如均匀采样、专家经验单点测试)以及当前的化工专用AI,都存在明显的局限——只能实现单一变量调整或局部参数优化,难以覆盖高维、非线性的复杂化学空间,导致资源浪费和效率低下。而化工AGI的自动

实验工具,能够支撑全流程自主优化:通过AI驱动的闭环逻辑,整合“需求解析-方案设计-实验执行-数据反馈-参数迭代”全环节,而不是局限在单一实验步骤上。在这一前沿领域,我国科研团队取得了具有引领性的进展,展现了从“全流程自主”到“人机自然交互”的多样化技术路径。一方面,以Zhu等^[53]研制的全流程“机器化学家”平台为代表,实现了实验范式的重要革新。该平台集成了移动机器人、智能化学工作站与云端数据库,其后台的“化工大脑”能够融合理论计算与机器学习,自主完成从智能文献阅读、实验方案设计、高通量合成表征到结果优化反馈的完整闭环。在一个典型案例中,面对超过55万种可能配比的高熵化合物催化剂筛选,该系统仅用5周时间就锁定了最优配方,把传统“穷举试错”方法所需的理论时间缩短了数个数量级。另一方面,实验工具的自主化正在朝着更灵活、智能的人机协同方向演进。Huang等^[54]开发了一种基于自然语言交互的AI驱动机器人合成平台,使研究人员能够直接使用自然语言描述合成目标,系统便可自动解析、规划并执行实验流程,这种全流程自主响应能力,正是化工AGI区别于化工专用AI的核心特征之一。

自动生产控制工具的核心定位,是依托人工智能实现大型工业装置实时优化与异常处置的全流程自主接管;对于大型化工生产环节而言,这也是化工AGI实现从传统静态数据监控,向全流程实时自主控制跨越的核心应用方向。帝国理工学院研究团队近期发表的工作,便针对这一应用方向提出了基于大语言模型(LLM)的自主工业控制代理框架,专门用于解决化工厂动态运行环境下的复杂故障处置问题^[55]。该框架搭建了包含操作执行、结果验证与重提示优化模块的多智能体(multi-agent)协同架构,无须依赖人工预先设定的规则体系,即可自主解析装置的实时生产数据,对生产过程中不可预见的工艺扰动完成自主响应与处置。通过与工厂数字孪生模型的持续实时交互与前置性策略验证,这类自主控制代理可实现快速的异常排查定位,同时直接生成并落地执行适配当前工况的最优操作调整指令。这类具备认知决策能力的工具落地应用,标志着人工智能技术正从研发实验室的理论验证场景,真正切入工业生产一线的分布式控制(DCS)系统,为大型化工装置赋予了异常自我修复、运行过程全自主优化的核心能力。

自动模拟、自动实验与自动生产控制形成协同闭环:模拟工具基于理论模型预测最优参数空间,为实验与生产提供靶向指导;实验与生产产生的真实数据被用来校准数字孪生与模拟模型,反哺精度提升;三者循环迭代,加速研发与工艺优化的进程。在新材料合成中,模拟先圈定温度、压力等参数的最优范围,实验在这个范围内精准探索并反馈数据,模拟再修正模型进一步缩小范围,最终经过数轮迭代确定最优工艺。在催化剂开发中,Zhong等^[56]利用主动机器学习方法加速了CO₂电催化剂的发现过程,通过高通量模拟筛选与实验验证相结合,大幅提升了催化剂筛选效率。在膜材料设计中,Budhathoki等^[57]通过高通量分子模拟系统筛选了MOF-聚合物混合基质膜,将计算预测的CO₂/N₂分离性能与实验数据对比验证,为膜材料的计算-实验协同设计提供了完整范式。这种模式融合了模拟的理论性与实验的真实性,大幅提升了研发效率。而三者的核心价值,在于让化工专用算力通过自主工具释放效能,实现算力按需分配、任务自主完成,不需要工程师介入烦琐操作。但需要注意的是,无论工具和算力多么先进,如果脱离了化工的

本质规律(如物理化学定律、质量守恒),AI仍然可能输出违背常识的结果。因此,对化工机理的理解仍然是底层支撑。那么,如何让AI真正理解化工过程的本质,确保决策的可靠性与可解释性?这就需要机理与数据的深度融合,构建化工AGI的认知根基。这种模式融合了模拟的理论性与实验的真实性,形成“模拟预测-实验验证-模型校准”的闭环,其认知工具体系如图4所示。

4 认知根基:机理与数据融合的化工AGI核心

数据、算力与工具构成了化工AGI的外在能力边界,而机理与数据的深度融合则决定了其内在认知深度。化工过程受严格的物理化学定律支配——质量守恒、能量守恒、动量守恒以及化学反应动力学共同构成了三传一反的理论基石。纯数据驱动模型虽然能在训练分布内实现高精度拟合,但一旦外推到未知工况,就可能生成违背基本物理化学原理的荒谬结果(如产物质量超过反应物、吸热反应自发升温)^[58]。因此,化工AGI的认知

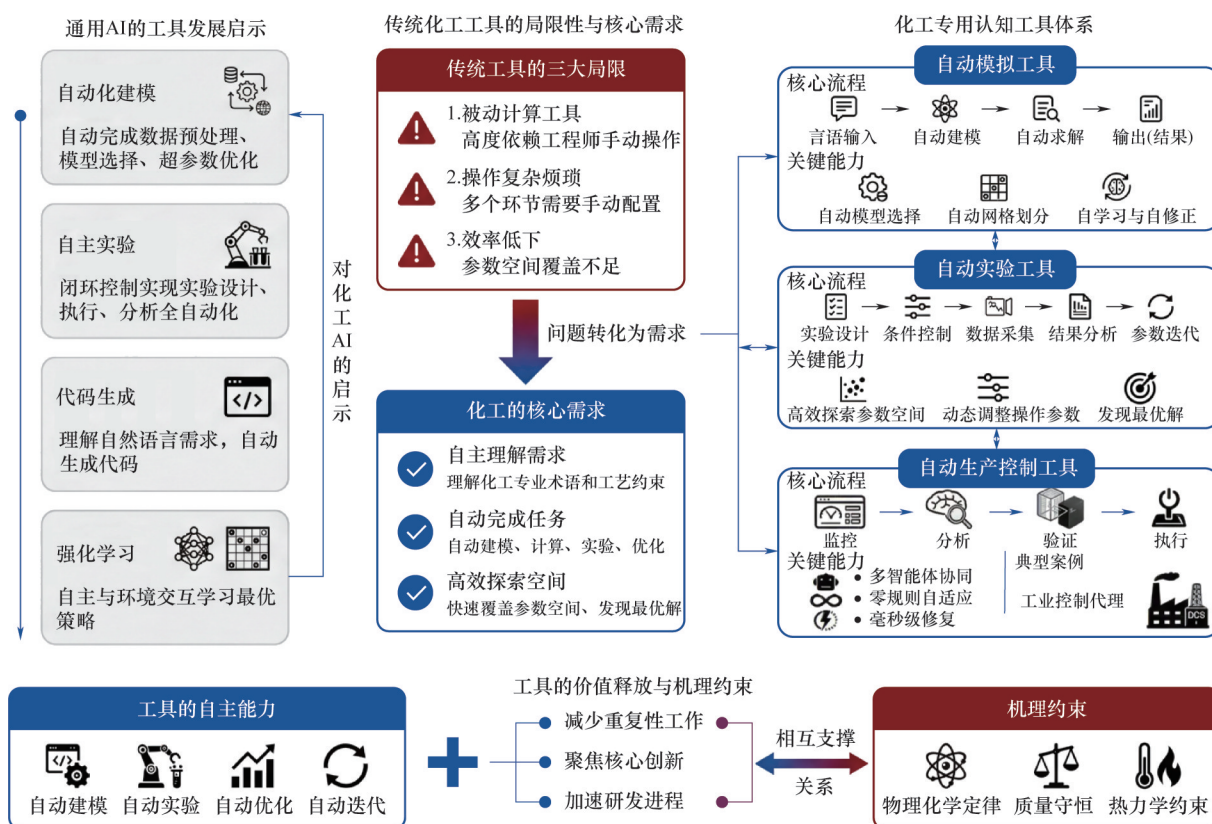


图4 化工AGI认知工具体系:自主模拟、实验与协同闭环

Fig.4 Cognitive tool system for chemical engineering AGI: autonomous simulation, experimentation, and collaborative closed loop

根基必须建立在机理知识与数据驱动的深度耦合之上,使模型不仅能拟合数据,更能理解化工过程的本质规律^[59]。化工生产涉及高温高压、易燃易爆、有毒有害等极端工况,对模型预测的可靠性与可解释性有着远超一般工业场景的严苛要求。然而,当前主流的深度学习模型本质上是高维非线性函数拟合器,其决策过程缺乏透明性,很难回答“为什么这样预测”的问题^[60]。

这种“黑箱”特性在化工领域引发了双重困境:外推失效困境——神经网络的泛化能力高度依赖训练数据的覆盖范围。当操作条件(温度、压力、流量、组分)超出历史数据边界时,模型预测可能严重偏离真实值。更危险的是,模型本身无法识别这种失效,仍然会输出看似合理的数值结果。在反应器温度预测中,如果训练数据只覆盖300~400 K范围,模型对450 K工况的预测可能完全失真,却不会给出任何警示。物理违背困境——纯数据驱动模型不具备物理化学常识,其输出可能违反最基本的守恒定律。在物料平衡预测中,模型可能给出进出口质量不守恒的结果;在反应动力学建模中,可能预测出活化能为负值或反应级数不合理的参数;在相平衡计算中,可能给出违反吉布斯相律的组成。这些物理上不可能的结果一旦用于过程控制或安全联锁,将带来灾难性后果。

Hinton^[61]在早期研究中发现:简单的神经网络通过学习词与词之间的统计关系就能涌现出语义表征,但模型底层并不真正理解语义。这种语义困惑映射到化工领域就是:模型能拟合历史数据中的输入-输出映射,却不理解这些映射背后的物理化学机制。化工生产涉及安全、环保、经济等重大责任,这种“知其然不知其所以然”的黑箱模型很难获得工程师与监管机构的信任^[62]。解决上述困境的根本出路,在于把人类积累的物理化学知识注入数据驱动模型,构建“机理约束+数据驱动”的混合建模范式。这一范式的核心价值体现在三个层面:约束性——机理知识为模型输出划定物理边界,确保预测结果不违反基本定律;解释性——模型参数或中间变量具有明确的物理意义(如速率常数、传递系数),便于工程师理解和验证;泛化性——物理规律的普适性使模型能更好地外推到训练数据未覆盖的工况。

值得注意的是,机理与数据的融合并非仅局限于模型结构优化或损失函数的针对性改造层面,从

数据输入端切入的机理导向特征提取,对于化工场景下高可靠性预测模型的构建,同样具有决定性意义。针对这一技术方向,中国石油大学(北京)张霖宙教授团队围绕复杂分子反应体系的过程放大难题,提供了具有工程参考价值的应用范例^[63]。他们提出了一套反应机理与深度迁移学习相融合的混合建模框架,其核心设计逻辑在于:首先通过构建分子层级的催化裂化反应动力学模型生成海量理论数据集,再基于该数据集完成分层神经网络的预训练,最终实现了对高度复杂的分子转化特征的精准提取。该工作更为核心的突破在于,将反应体系的底层物性信息作为机理先验特征,直接嵌入神经网络的输入端,这一设计有效克服了中试放大阶段高质量数据稀缺的瓶颈,仅需通过极少量的现场真实数据完成模型微调,即可实现中试装置产物分布的精准计算,同步完成工艺参数的多目标优化。这一完整的工业应用案例清晰印证,将物理化学机理前置到数据预处理与特征提取环节,是全面提升化工大模型泛化能力、保障其工业场景落地可行性的核心支撑。机理与数据的融合不是简单的“机理模型+数据模型”串并联,而是需要根据融合深度构建层次化的架构。在夯实了输入端的机理特征映射这一基础之后,将视野深入到AI模型内部的运算逻辑时,可以进一步将机理与数据的融合划分为循序渐进的三重架构,具体如表4所示。

规律约束层的核心思想是把物理化学定律作为软约束引入模型训练过程。具体实现上,在损失函数中除了预测误差项,还额外添加物理规律违背惩罚项:

$$L = L_{\text{data}} + \lambda L_{\text{physics}} \quad (1)$$

式中, L_{data} 为数据拟合误差; L_{physics} 为物理约束违背程度; λ 为平衡系数。

在化工过程建模中,典型的物理约束包括:质量守恒——反应器进出口各组分的质量平衡;能量守恒——热量输入、输出与累积的平衡;热力学一致性——相平衡计算需满足等温等压条件下各相化学势相等;反应化学计量——产物生成量与反应物消耗量需符合化学计量关系。Nascimento等^[64]在锂离子电池建模研究中,将电化学机理(Nernst方程、Butler-Volmer方程)作为软约束嵌入混合物理信息神经网络。该模型不仅在闭环实验中发现了延长电池寿命的最优充电策略,更重要的是在未见过的电解质体系中表现出了良好的泛化能力。Wu

表 4 机理-数据融合的三重架构

Table 4 Three-level architecture for mechanism-data integration

融合层次	机理角色	技术实现	化工典型应用	优势与局限
规律约束层	外部校验器:在损失函数中惩罚违反物理规律的预测	约束损失函数、拉格朗日乘子法、软/硬约束正则化	精馏塔物料平衡、反应器能量守恒、相平衡热力学一致性	实现简单、通用性强;但机理仅作为惩罚项,未融入特征学习
结构嵌入层	先验骨架:将控制方程的数学形式编码为网络结构	PINNs、神经 ODE、反应动力学神经网络	流场速度-压力耦合求解、反应路径与速率常数估计、多物理场耦合模拟	网络权重具有物理意义;但需针对具体方程设计专用架构
知识发现层	进化对象:从数据中自动发现或修正机理模型	符号回归、SINDy 稀疏识别、遗传编程	新型传递系数关联式、催化活性描述符、反应网络拓扑推断	可突破现有机理局限;但对数据质量要求高、结果需实验验证

等^[65]将物料与能量平衡方程作为约束加入 LSTM 网络,用于预测精馏塔产品浓度。在进料组成波动、回流比调整等工况变动场景下,预测鲁棒性比纯数据模型提升了约 35%。规律约束层的优势在于实现简单、适用范围广,不需要修改网络架构就能把物理知识注入任意深度学习模型。但其局限性也很明显:机理知识只是作为外部惩罚存在,并没有真正融入模型的特征提取与表示学习过程,模型内部仍然是“黑箱”。

结构嵌入层的核心思想是把物理方程的数学形式直接编码为网络结构,使网络架构与物理过程形成同构映射。这一策略的哲学基础在于:与其让网络从数据中重新发现物理规律,不如把已经验证的规律预先植入网络骨架,让数据学习聚焦于规律参数化与复杂耦合效应的捕捉。物理信息神经网络是这一思想的典范框架^[66],简单来说,就是为神经网络“植入物理常识”,使其在学习数据规律时必须遵守物理定律(如守恒律)。PINNs 把偏微分方程的残差直接引入损失函数:对于控制方程 $\mathcal{N}[u] = 0$ (如 Navier-Stokes 方程),网络在最小化边界/初始条件误差的同时,被强制要求其输出 u_{NN} 满足 $\mathcal{N}[u_{NN}] \approx 0$ 。这一设计使网络能够在极稀疏的监督数据下求解正问题(给定参数求解场分布),或者在给定部分观测的条件下求解逆问题(反演未知参数)。在化工领域,PINNs 已成功应用于管内湍流换热模拟、多孔介质渗流求解、反应器浓度场重构等场景。更深层次的结构嵌入体现在反应动力学专用网络的设计上。Ji 等^[67]开发的化学反应神经网络把质量作用定律与 Arrhenius 方程直接构建在网络结构中,通俗而言,就是让神经网络的架构本身契合化学反应的基本规律,无须额外施加约束。每个网络层对应一个基元反应步骤,层间权重对应速率常数 k_j ,激活函数被设计为温度的 Arrhenius 形式

$\exp(-E_a/RT)$ 。这种设计确保了:(1)网络前向传播在数学上等价于求解反应动力学 ODE;(2)训练得到的权重(活化能 E_a 、指前因子 A)天然具有物理意义;(3)网络能从时间序列浓度数据中推断反应路径拓扑。结构嵌入层的核心价值在于实现了“白箱化”,网络参数不再是没物理意义的数值,而是可解释、可验证的物理量。其局限在于需要针对不同物理过程设计专用架构,通用性不如规律约束层。

知识发现层代表机理-数据融合的较高境界:不仅用机理约束数据,还让 AI 从数据中发现新机理、修正旧机理,实现知识的自主进化。这一层次的价值在于突破现有机理模型的局限——化工领域仍有大量过程(复杂反应的副反应路径、湍流多相流的传递系数、催化剂失活机制)尚未被充分理解,传统机理模型依赖简化假设,而海量模拟与实验数据中可能隐藏着更精准的规律^[68]。知识发现层的实现高度依赖前文所述的数据自主造血能力与专用算力支撑。之前提出的多尺度模拟数据范式(分子动力学、粗粒化模拟、CFD 高通量计算)为知识发现提供了“矿藏”;随后论述的化工专用超算与加速芯片则为大规模数据挖掘提供了算力保障。三者形成闭环:算力驱动模拟→模拟生成数据→数据反哺机理→机理指导模拟。机器学习势能是知识发现的典型范例,具体来说,就是通过 AI 学习原子间的相互作用规律,同时兼顾量子力学的高精度与经典力场的高效率。传统分子动力学模拟面临精度与效率的两难:量子力学方法(如 DFT)精度高但计算量巨大,经典力场方法效率高但精度受限于参数化形式。MLP 通过神经网络学习原子构型与势能面的映射关系,在保持接近 DFT 精度的同时把计算效率提升了 3~5 个数量级。更重要的是,MLP 的训练过程本身就是一种机理发现——网络从第一性原理计算数据中自动提取原子间相互作用的

复杂模式,这些模式可能揭示传统解析势函数未能捕捉的物理效应。在化工领域,MLP已被用于预测催化剂表面吸附能、离子液体热力学性质、高分子链构象能量等,为分子尺度机理的精细化提供了新工具^[69]。符号回归与稀疏识别提供了从数据到解析表达式的直接通路,也就是让AI自主推导公式,不需要预先设定函数形式。与传统回归拟合预设函数形式的参数不同,符号回归从候选算子库(加减乘除、指数对数、三角函数等)中自动搜索最能描述数据的数学表达式。化工AGI的一个可行场景:通过高通量CFD模拟生成数万组多相流反应器的操作参数-传递系数数据,再利用符号回归从中挖掘新型曳力模型或传质系数关联式。这种“计算实验+智能发现”的范式有望产生比传统经验公式(如Ergun方程、Ranz-Marshall关联式)更精准、适用范围更广的新机理表达。反应网络自动推断是知识发现在化学反应领域的前沿方向。复杂化工过程(如催化裂化、78煤气化、生物质转化)涉及成百上千种中间物种与反应路径,传统方法很难穷举。结合高通量量子化学计算与图神经网络,有望实现从反应物-产物数据自动推断反应网络拓扑、识别主反应与副反应路径、预测中间体稳定性。这一能力对于反应器设计、选择性优化、副产物控制具有重要的工程价值。守恒律与不变量发现代表知识发现的更高层次。复杂化工系统是否存在尚未被认识的守恒量或对称性?神经网络能否从过程数据中自动发现这些隐藏的物理结构?这些问题目前还没有成熟答案,但为化工AGI的长远发展指明了方向。需要强调的是,知识发现层的成果必须经过严格的实验验证与物理审查。AI发现的新机理可能只是数据噪声或局部相关性的伪规律,必须通过独立实验数据验证其普适性,用量纲分析、极限行为分析等手段检验其物理合理性。知识发现层的价值不在于替代人类的机理研究,而在于为化工科学家提供“假说生成器”——从海量数据中筛选出值得深入研究的候选规律,加速科学发现的循环。

三传一反是化工过程的核心机理框架,也是机理-数据融合架构落地的具体载体^[5]。动量传递方面,Navier-Stokes方程是流体流动的基本控制方程。在规律约束层,可以通过质量守恒(连续性方程)与动量守恒约束流场预测模型,确保预测的速度场没有质量“凭空产生或消失”;在结构嵌入层,PINNs可以把N-S方程残差作为损失项,实现从稀疏PIV/

LDV测量数据到全流场的重构,这对于工业反应器内部流场诊断具有重要价值;在知识发现层,结合前述化工专用超算进行大规模湍流DNS模拟,再利用稀疏识别算法从模拟数据中提取低维湍流模型,有望发现超越传统RANS模型(如 $k-\varepsilon$ 、 $k-\omega$)的新型封闭关系。热量传递方面,傅里叶定律与能量守恒方程描述热量传递过程。在规律约束层,换热器模型可以嵌入热量平衡约束($Q_{\text{hot}} = Q_{\text{cold}}$),避免预测出热量不守恒的结果;在结构嵌入层,PINNs可以耦合能量方程与动量方程,求解对流-导热耦合问题,用于反应器热点预测与温度场重构;在知识发现层,通过高通量CFD模拟生成不同几何结构、流动工况下的传热数据,再利用符号回归发现新型Nusselt数关联式,这些数据驱动的关联式可能比传统Dittus-Boelter、Sieder-Tate等经验公式具有更宽的适用范围。质量传递方面,菲克定律与组分守恒方程描述传质过程。在规律约束层,吸收塔、萃取塔模型可以嵌入各组分物料平衡约束;在结构嵌入层,可以把双膜理论或渗透理论的数学形式编码为网络层,使传质系数 k_L 、 k_G 等参数具有明确的物理意义;在知识发现层,结合前述分子动力学模拟数据(如扩散系数、溶解度的分子层面预测),利用机器学习建立分子结构-传质性能的定量构效关系,为新型分离介质(膜材料、吸附剂、萃取剂)的理性设计提供指导。化学反应方面,反应动力学方程描述反应速率与转化规律。在规律约束层,反应器模型可以嵌入原子守恒(C、H、O等元素的进出平衡)与热力学可行性约束(反应方向需符合 ΔG 判据);在结构嵌入层,化学反应神经网络把质量作用定律与Arrhenius方程编码为网络结构,使活化能、指前因子等动力学参数可以直接从网络权重读出;在知识发现层,机器学习势能发挥关键作用——通过高精度量子化学计算(DFT/CCSD(T))生成势能面数据,训练MLP后可以接近第一性原理的精度、经典力场的效率进行反应路径搜索与过渡态定位,结合前述化工专用加速芯片,有望实现复杂催化反应网络的自动推断,发现传统实验难以识别的中间体与副反应路径。

三重融合架构与前文数据、算力、工具三大支柱的协同关系可概括为:数据层提供“原料”(多尺度模拟数据+实验数据)→算力层提供“引擎”(超算/专用芯片加速模拟与训练)→机理层提供“炼化能力”(从原始数据中提炼物理规律)→工具层实现

“落地”(将融合模型部署于自动模拟/自动实验平台)。这种四位一体的协同确保了化工AGI的可信性(规律约束层保障物理合规)、可解释性(结构嵌入层赋予参数物理意义)与可进化性(知识发现层推动机理迭代),从根本上解决了纯数据驱动模型的信任危机,构筑起化工AGI区别于通用AI的核心认知能力。四重融合架构与前文数据、算力、工具三大支柱的协同关系,以及在三传一反中的落地路径,如图5所示。

5 结论与展望

构建化工领域通用人工智能(化工AGI)是化工行业智能化转型的核心目标,其核心逻辑是构建数据-算力引擎-工具-机理四大支柱的协同体系,这四大支柱以化工为本,深度适配三传一反的领域特性,形成自我强化的生态闭环:数据作为燃料,通过多尺度模拟自主造血+实验校准的模式,破解化工数据稀缺难题;算力作为引擎,以专为化工计算设计的芯片与硬件架构为核心,提供高效计算支撑;工具作为载体,通过自动模拟、自动实验以及自动生产控制实现AI对化工研发与生产的自主服务;机理作为根基,通过层次化融合确保AI模型的可信性与可解释性。四大支柱并非技术的简单堆砌,而是

围绕化工行业的实际需求,实现通用AI技术与化工专业知识的深度适配与创新。

当前,化工AGI的构建仍然面临多重关键挑战,这些挑战进一步印证了“数据需要经过机理加工才能成为工业级可用信息”这一核心认知:第一,多尺度数据的标准化与融合问题——核心是解决数据无序化的难题,为工业数据的规范化表达奠定基础;第二,化工专用芯片的研发成本控制与技术门槛突破——需要为机理知识的解析、耦合与应用提供稳定高效的算力支撑;第三,自动工具对复杂化工场景的适配性提升——需要实现从原始数据到化工领域专业知识的自动化转化;第四,面向深层机理发现的AI算法创新——需要推动模型实现对数据背后隐性规律的自主挖掘^[2]。解决上述问题离不开跨学科的深度协作:化工工程师聚焦机理知识梳理与工业场景需求拆解,为AI模型的工业适配提供核心依据;计算机科学家致力于算法优化与硬件架构升级,打造高效的数据机理化加工工具;材料科学家与数学家则专注于机理模型的完善与优化,为数据的机理化解析提供坚实的理论框架。未来化工AGI的关键突破方向可以聚焦于以下维度:一是构建统一的化工数据共享平台,实现多源异构数据的整合与高效利用;二是开发低成本、高性能的化工专用芯片,降低数据机理化加工过程中的算力

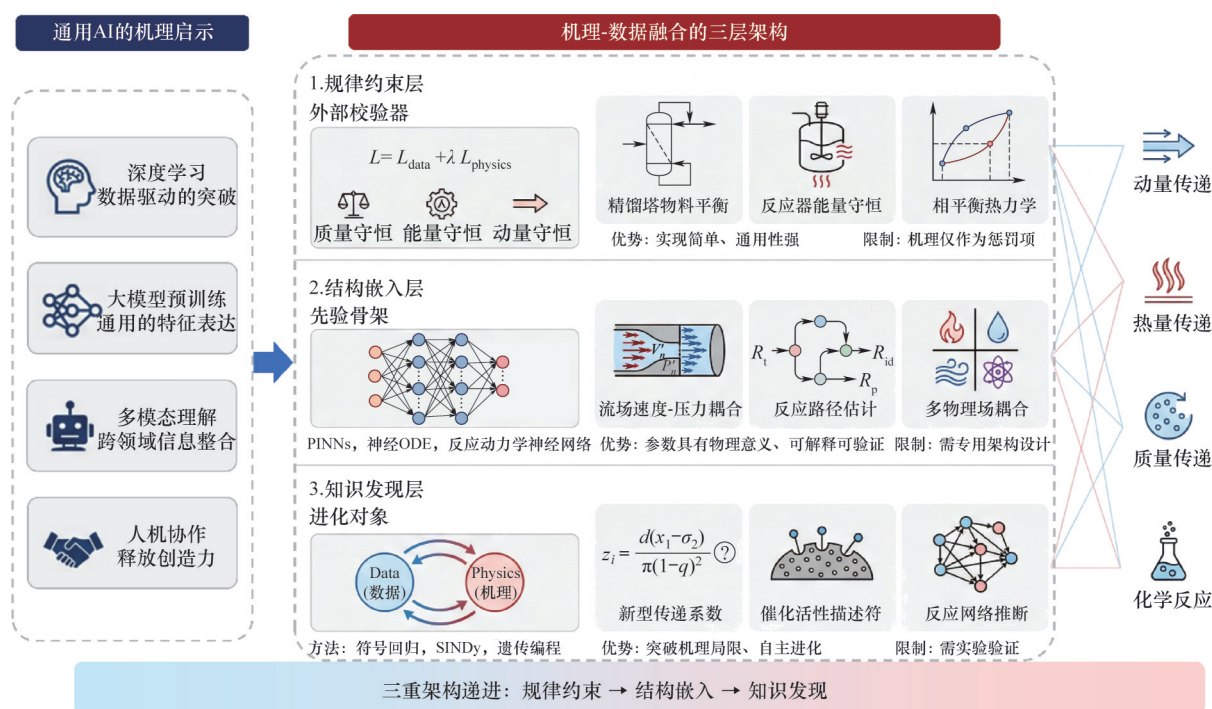


图5 化工机理与数据融合框架

Fig.5 Framework for integrating mechanisms and data in chemical engineering

消耗;三是构建融合数字孪生与多智能体的自主工具体系,提升对复杂多相流、强耦合反应体系及全厂级自主控制等典型化工场景的适配能力;四是研发更高效的机理发现算法,强化模型对数据中隐性规律的自主提取能力。通过这些突破,最终实现化工AGI从数据积累到智能生成的关键跨越^[70]。

展望未来,化工AGI的崛起将重塑化工工程师的角色定位,同时也需要警惕AI技术普遍面临的人才成长阶梯断裂风险:传统化工工程师需要通过数据整理、手动建模、反复实验等基础实践来积累隐性知识,比如对工艺参数波动的直觉、对异常数据的判断力;而化工AGI对重复性任务的替代,可能让新一代工程师失去试错成长的机会,导致未来资深专家的断代危机。因此,化工领域的人机协同不能只停留在效率提升上,还需要构建适配化工AGI的新型学徒制:一方面,未来的化工工程师将转变为系统架构师与创新战略家,核心价值聚焦于三大任务——把模糊的工程目标转化为AI可以理解的具体问题、基于安全环保要求划定AI的优化边界、评估AI决策的物理合理性;另一方面,需要通过场景化训练保留成长路径,例如在教育阶段设置没有AI辅助的基础实验项目,在工作中增加AI结果审查与修正任务,让工程师在与化工AGI的互动中,既掌握智能工具的应用逻辑,又积累化工领域的隐性知识^[71]。这种效率与传承并重的人机协同,才是化工行业智能化转型的可持续形态:AI作为能力放大器延伸工程师的创新边界,而工程师通过批判性实践确保化工专业智慧的代际传递,避免陷入依赖AI却切断人才供给的困境。最终,化工AGI的构建必须回归“化工为本”的核心理念——所有技术创新都应该服务于化工产业的实际需求,而不是追求技术的炫技;所有模型与工具都应该适配三传一反的本质规律,而不是脱离化工场景空谈智能。只有这样,AI才能真正成为推动化工行业向更高效、更绿色、更安全方向发展的强大引擎,为解决能源短缺、环境污染、资源浪费等全球性挑战贡献化工力量。化工AGI的未来,不是AI替代化工工程师,而是AI与化工工程师协同进化,共同开启化工行业的智能化新篇章^[72]。

参考文献

- [1] 曾玉娇,肖旸,杨刚,等. 基于机理与数据混合驱动的湿法磷酸生产过程代理建模与优化[J]. 化工学报, 2024, 75(3): 936-944.
Zeng Y J, Xiao X, Yang G, et al. Surrogate modeling and optimization of wet phosphoric acid production process based on mechanism and data hybrid driven[J]. CIESC Journal, 2024, 75(3): 936-944.
- [2] 吉远辉,朱家华,穆立文,等. 化工基础数据获取新范式: 机制+数据驱动[J]. 中国科学基金, 2024, 38(4): 712-718.
Ji Y H, Zhu J H, Mu L W, et al. New Paradigm for determining fundamental chemical engineering data: a “mechanism+data driven” strategy[J]. Bulletin of National Natural Science Foundation of China, 2024, 38(4): 712-718.
- [3] 陈宇翔,刘伟磊,龚子君,等. 机器学习辅助乙硫醇高效吸收溶剂分子设计[J]. 化工学报, 2024, 75(3): 914-923.
Chen Y X, Liu C L, Gong Z J, et al. Machine learning-assisted solvent molecule design for efficient absorption of ethanethiol[J]. CIESC Journal, 2024, 75(3): 914-923.
- [4] 刘东飞,张帆,刘铮,等. 机器学习势及其在分子模拟中的应用综述[J]. 化工学报, 2024, 75(4): 1241-1255.
Liu D F, Zhang F, Liu Z, et al. A review of machine learning potentials and their applications to molecular simulation[J]. CIESC Journal, 2024, 75(4): 1241-1255.
- [5] 王凯,袁志宏,王笑楠,等. 化学品智能制造的科学基础[J]. 中国科学基金, 2023, 37(5): 811-817.
Wang K, Yuan Z H, Wang X N, et al. Scientific fundamentals of artificial intelligence involved chemical manufacturing[J]. Bulletin of National Natural Science Foundation of China, 2023, 37(5): 811-817.
- [6] 缪青海,王兴霞,杨静,等. 从基础智能到通用智能: 基于大模型的GenAI和AGI之现状与展望[J]. 自动化学报, 2024, 50(4): 674-687.
Miao Q H, Wang X X, Yang J, et al. From foundation intelligence to general intelligence: the state-of-art and perspectives of GenAI and AGI based on foundation models[J]. Acta Automatica Sinica, 2024, 50(4): 674-687.
- [7] 赵志君,庄馨予. 中国人工智能高质量发展: 现状、问题与方略[J]. 改革, 2023(9): 11-20.
Zhao Z J, Zhuang X Y. High-quality development of artificial intelligence in China: current situation, problems and strategies [J]. Reform, 2023(9): 11-20.
- [8] Schweidtmann A M, Esche E, Fischer A, et al. Machine learning in chemical engineering: a perspective[J]. Chemie Ingenieur Technik, 2021, 93(12): 2029-2039.
- [9] Herres-Pawlis S, Bach F, Bruno I J, et al. Minimum information standards in chemistry: a call for better research data management practices[J]. Angewandte Chemie International Edition, 2022, 61(51): e202203038.
- [10] Willoughby C, Frey J G. Data management matters[J]. Digital Discovery, 2022, 1(3): 183-194.
- [11] 李梓良,张玮,胡恒,等. 基于生成对抗网络的化工过程低成本mGAN-NN建模[J]. 化工进展, 2025, 44(4): 1978-1986.
Li Z L, Zhang W, Hu H, et al. mGAN-NN method for low-cost chemical process modeling based on generative adversarial networks[J]. Chemical Industry and Engineering Progress, 2025, 44(4): 1978-1986.
- [12] Eibeck A, Lim M Q, Kraft M. J-Park Simulator: an ontology-based platform for cross-domain scenarios in process industry[J]. Computers & Chemical Engineering, 2019, 131: 106586.
- [13] Jiang Z H, Xu C, Liu J S, et al. A dual closed-loop digital twin construction method for optimizing the copper disc casting process

- [J]. IEEE/CAA Journal of Automatica Sinica, 2024, **11**(3): 581–594.
- [14] 郭洪霞. 高分子粗粒化分子动力学模拟进展[J]. 高分子通报, 2011(10): 154–163.
- Guo H X. Progress on coarse-grained molecular dynamics simulation of polymers[J]. Polymer Bulletin, 2011(10): 154–163.
- [15] Marrink S J, Risselada H J, Yefimov S, et al. The MARTINI force field: coarse grained model for biomolecular simulations[J]. The Journal of Physical Chemistry B, 2007, **111**(27): 7812–7824.
- [16] Wu Z H, Zhou T H. Structural coarse-graining *via* multiobjective optimization with differentiable simulation[J]. Journal of Chemical Theory and Computation, 2024, **20**(6): 2605–2617.
- [17] Noid W G. Perspective: advances, challenges, and insight for predictive coarse-grained models[J]. The Journal of Physical Chemistry B, 2023, **127**(19): 4174–4207.
- [18] Li J H. The principle of compromise-in-competition: understanding mesoscale complexity of different levels[J]. Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences, 2024, **480**(2301): 20240031.
- [19] Guo L, Meng F Y, Qin P F, et al. A case study applying mesoscience to deep learning[J]. Engineering, 2024, **39**: 84–93.
- [20] Jain A, Hautier G, Moore C J, et al. A high-throughput infrastructure for density functional theory calculations[J]. Computational Materials Science, 2011, **50**(8): 2295–2310.
- [21] Horton M K, Huck P, Yang R X, et al. Accelerated data-driven materials science with the materials project[J]. Nature Materials, 2025, **24**(10): 1522–1532.
- [22] Jain A, Ong S P, Hautier G, et al. Commentary: the materials project: a materials genome approach to accelerating materials innovation[J]. APL Materials, 2013, **1**: 011002.
- [23] Zhu J X, Cheng J. Machine learning potential for electrochemical interfaces with hybrid representation of dielectric response[J]. Physical Review Letters, 2025, **135**: 018003.
- [24] Wang M, Ishii A, Sakaushi K. Accelerated electrocatalyst degradation testing by accurate and robust forecasting of multidimensional kinetic model with Bayesian data assimilation [J]. ACS Energy Letters, 2025, **10**(1): 22–29.
- [25] Zhao L F, Li Z, Wang Z C, et al. Active- and transfer-learning applied to microscale-macroscale coupling to simulate viscoelastic flows[J]. Journal of Computational Physics, 2021, **427**: 110069.
- [26] Shahnam M, Gel A, Subramanian A K, et al. Uncertainty quantification analysis of both experimental and CFD simulation data of a bench-scale fluidized bed gasifier[R]. Morgantown: National Energy Technology Laboratory (NETL), Research and Innovation Center, 2017.
- [27] Kusne A G, Yu H S, Wu C M, et al. On-the-fly closed-loop materials discovery *via* Bayesian active learning[J]. Nature Communications, 2020, **11**: 5966.
- [28] Steinmann S N, Hermawan A, Bin Jassar M, et al. Autonomous high-throughput computations in catalysis[J]. Chem Catalysis, 2022, **2**(5): 940–956.
- [29] Tom G, Schmid S P, Baird S G, et al. Self-driving laboratories for chemistry and materials science[J]. Chemical Reviews, 2024, **124**(16): 9633–9732.
- [30] Tran K, Ulissi Z W. Active learning across intermetallics to guide discovery of electrocatalysts for CO₂ reduction and H₂ evolution[J]. Nature Catalysis, 2018, **1**: 696–703.
- [31] Mu Y, Li N, Hang L Q, et al. Investigation of the rheological behaviors of polymeric materials in the film casting process through multiscale modeling and simulation method[J]. Macromolecular Theory and Simulations, 2019, **28**(3): 1900001.
- [32] 颜同安. 数据驱动下用于碳捕集纳米多孔材料的筛选与设计[D]. 北京: 北京化工大学, 2023.
- Yan T A. Data-driven screening and design of nanoporous materials for carbon capture[D]. Beijing: Beijing University of Chemical Technology, 2023.
- [33] Mao R Z, Zhang M, Wang Y R, et al. An integrated framework for accelerating reactive flow simulation using GPU and machine learning models[J]. Proceedings of the Combustion Institute, 2024, **40**(1/2/3/4): 105512.
- [34] Mittal S, Vetter J S. A survey of CPU-GPU heterogeneous computing techniques[J]. ACM Computing Surveys, 2015, **47**(4): 1–35.
- [35] Horowitz M. 1.1 Computing's energy problem (and what we can do about it) [C]//2014 IEEE International Solid-State Circuits Conference Digest of Technical Papers (ISSCC). IEEE, 2014.
- [36] Shaw D E, Deneroff M M, Dror R O, et al. Anton, a special-purpose machine for molecular dynamics simulation[J]. Communications of the ACM, 2008, **51**(7): 91–97.
- [37] Ohmura I, Morimoto G, Ohno Y, et al. MDGRAPE-4: a special-purpose computer system for molecular dynamics simulations[J]. Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences, 2014, **372**(2021): 20130387.
- [38] Mayr C, Hoepfner S, Furber S. SpiNNaker 2: A 10 million core processor system for brain simulation and machine learning[EB/OL]. 2019, arXiv: 1911.02385. <https://arxiv.org/abs/1911.02385>.
- [39] Joubert W, Archibald R, Berrill M, et al. Accelerated application development: the ORNL Titan experience[J]. Computers & Electrical Engineering, 2015, **46**: 123–138.
- [40] Santos K, Moore S, Oppelstrup T, et al. Breaking the molecular dynamics timescale barrier using a wafer-scale system[C]//SC24: International Conference for High Performance Computing, Networking, Storage and Analysis. IEEE, 2024.
- [41] Hamm P. Toward an FPGA-based dedicated computer for molecular dynamics simulations[J]. The Journal of Chemical Physics, 2025, **162**(5): 054108.
- [42] 徐硕, 毕可鑫, 邱彤, 等. 智能过程系统工程(iPSE): 当过程系统工程与新一代人工智能相遇[J]. 化工进展, 2026, **45**(4): 1877–1899.
- Xu S, Bi K X, Qiu T, et al. Intelligent process systems engineering (iPSE): when process systems engineering meets new generation artificial intelligence[J]. Chemical Industry and Engineering Progress, 2026, **45**(4): 1877–1899.
- [43] Trotter J D, Langguth J, Cai X. Targeting performance and user-friendliness: GPU-accelerated finite element computation with automated code generation in FEniCS[J]. Parallel Computing, 2023, **118**: 103051.
- [44] Porcu R, Musser J, Almgren A S, et al. MFix-Exa: CFD-DEM simulations of thermodynamics and chemical reactions in multiphase flows[J]. Chemical Engineering Science, 2023, **273**: 118614.
- [45] Robertson L A, Shkrob I A, Lewis R, et al. High-throughput discovery illuminates design principles and limits for long-lived

- charged species in organic electrolytes[J]. *Journal of the American Chemical Society*, 2025, **147**(41): 37211–37222.
- [46] Delgado-Licona F, Alsaia A, Dickerson H, et al. Flow-driven data intensification to accelerate autonomous inorganic materials discovery[J]. *Nature Chemical Engineering*, 2025, **2**(7): 436–446.
- [47] Boiko D A, MacKnight R, Kline B, et al. Autonomous chemical research with large language models[J]. *Nature*, 2023, **624**(7992): 570–578.
- [48] Du W L, Yang S Y. The potential and challenges of large language model agent systems in chemical process simulation: from automated modeling to intelligent design[J]. *Frontiers of Chemical Science and Engineering*, 2025, **19**(10): 99.
- [49] Ren L, Wang H T, Wang Y Q, et al. Foundation models for the process industry: challenges and opportunities[J]. *Engineering*, 2025, **52**: 53–59.
- [50] 鹿兰婷, 康胜, 许文珂, 等. 化工领域中的人工智能: 人工神经网络技术的应用与前景[J]. *化工进展*, 2025, **44**(8): 4808–4820.
- Lu L T, Kang S, Xu W K, et al. Artificial intelligence in the chemical industry: applications and prospects of artificial neural network technology[J]. *Chemical Industry and Engineering Progress*, 2025, **44**(8): 4808–4820.
- [51] LeCun Y. A path towards autonomous machine intelligence version 0.9.2, 2022–06–27[J]. *OpenReview*, 2022, **62**(1): 1–62.
- [52] 侯卫锋, 古绍武, 张志铭, 等. 工业大模型赋能的新型流程工业智能工厂核心工业软件体系[J]. *中国科学(信息科学)*, 2025, **55**(7): 1783–1800.
- Hou W F, Gu S W, Zhang Z M, et al. Core industrial software system of the new process industrial intelligent factory enabled by the industrial large model[J]. *Scientia Sinica (Informationis)*, 2025, **55**(7): 1783–1800.
- [53] Zhu Q, Zhang F, Huang Y, et al. An all-round AI-chemist with a scientific mind[J]. *National Science Review*, 2022, **9**(10): nwac190.
- [54] Huang L, Zhang C, Fu Y, et al. Natural-language-interfaced robotic synthesis for AI-copilot-assisted exploration of inorganic materials[J]. *Journal of the American Chemical Society*, 2025, **147**(26): 23014–23025.
- [55] Vyas J, Mercangöz M. Autonomous industrial control using an agentic framework with large language models[J]. *IFAC-PapersOnLine*, 2025, **59**(6): 349–354.
- [56] Zhong M, Tran K, Min Y M, et al. Accelerated discovery of CO₂ electrocatalysts using active machine learning[J]. *Nature*, 2020, **581**(7807): 178–183.
- [57] Budhathoki S, Ajayi O, Steckel J A, et al. High-throughput computational prediction of the cost of carbon capture using mixed matrix membranes[J]. *Energy & Environmental Science*, 2019, **12**(4): 1255–1264.
- [58] Karpatne A, Atluri G, Faghmous J H, et al. Theory-guided data science: a new paradigm for scientific discovery from data[J]. *IEEE Transactions on Knowledge and Data Engineering*, 2017, **29**(10): 2318–2331.
- [59] Psychogios D C, Ungar L H. A hybrid neural network-first principles approach to process modeling[J]. *AIChE Journal*, 1992, **38**(10): 1499–1511.
- [60] Gunning D, Stefik M, Choi J, et al. XAI—explainable artificial intelligence[J]. *Science Robotics*, 2019, **4**(37): eaay7120.
- [61] Hinton G E. Learning distributed representations of concepts[C]// *Proceedings of the Eighth Annual Conference of the Cognitive Science Society*. Amherst, MA, 1986: 12.
- [62] Rudin C. Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead[J]. *Nature Machine Intelligence*, 2019, **1**(5): 206–215.
- [63] Chen Z Y, Xie Y Q, Xu C M, et al. Scale-up of complex molecular reaction system by hybrid mechanistic modeling and deep transfer learning[J]. *Nature Communications*, 2025, **16**: 8905.
- [64] Nascimento R G, Corbetta M, Kulkarni C S, et al. Hybrid physics-informed neural networks for lithium-ion battery modeling and prognosis[J]. *Journal of Power Sources*, 2021, **513**: 230526.
- [65] Alhajeri M S, Abdullah F, Wu Z, et al. Physics-informed machine learning modeling for predictive control using noisy data[J]. *Chemical Engineering Research and Design*, 2022, **186**: 34–49.
- [66] Raissi M, Perdikaris P, Karniadakis G E. Physics-informed neural networks: a deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations[J]. *Journal of Computational Physics*, 2019, **378**: 686–707.
- [67] Ji W Q, Deng S L. Autonomous discovery of unknown reaction pathways from data by chemical reaction neural network[J]. *The Journal of Physical Chemistry A*, 2021, **125**(4): 1082–1092.
- [68] Ulissi Z W, Tang M T, Xiao J P, et al. Machine-learning methods enable exhaustive searches for active bimetallic facets and reveal active site motifs for CO₂ reduction[J]. *ACS Catalysis*, 2017, **7**(10): 6600–6608.
- [69] Behler J, Parrinello M. Generalized neural-network representation of high-dimensional potential-energy surfaces[J]. *Physical Review Letters*, 2007, **98**(14): 146401.
- [70] 刘科均. 浅谈人工智能技术在化工生产自动化控制系统中的应用[J]. *石化技术*, 2022, **29**(6): 261–263.
- Liu K J. Application of artificial intelligence technology in automatic control system of chemical production[J]. *Petrochemical Industry Technology*, 2022, **29**(6): 261–263.
- [71] 吉远辉, 苏隽奕, 李弋, 等. 人工智能背景下化工专业课程教学改革研究与实践[J]. *化工高等教育*, 2024, **41**(3): 9–13, 105.
- Ji Y H, Su J Y, Li Y, et al. Research and practice on courses teaching reform in chemical engineering under the background of artificial intelligence[J]. *Higher Education in Chemical Engineering*, 2024, **41**(3): 9–13, 105.
- [72] 陈光宇, 杨欣昱, 梁娜, 等. 以人为本的人工智能工程伦理准则探析[J]. *电子科技大学学报(社科版)*, 2020, **22**(6): 32–38.
- Chen G Y, Yang X Y, Liang N, et al. Analysis of people-oriented engineering ethics principles for artificial intelligence[J]. *Journal of University of Electronic Science and Technology of China (Social Sciences Edition)*, 2020, **22**(6): 32–38.