

AI+材料科学：大模型在新材料领域的研究进展与展望

张林桐¹, 刘传宝², 白 洋²

(1. 北京科技大学自然科学基础实验中心, 北京 100083; 2. 北京科技大学新材料技术研究院, 北京 100083)

摘 要: 随着数据驱动科学研究第四范式的深入发展, 以大模型为代表的人工智能技术正深度重构材料研发模式。本文系统综述了材料大模型在晶体结构设计与发现、性能预测与基础知识挖掘、具体材料垂类大模型、冶金工艺优化及多智能体协同框架的应用场景。在晶体结构设计与发现方面, GNoME、CrystalLLM 分别通过主动学习与自回归建模, 显著提升了稳定新材料的发现速率; 在性能预测与基础知识挖掘领域, MatChat 与 Galactica 等利用检索增强生成与大规模科学语料预训练, 实现了从海量非结构化文献中精准提取规律并预测材料属性; 针对具体材料垂类大模型, PolyLLMem、MatMind 及 ChipNeMo 等通过整合高分子、无机非金属及半导体设计的专业先验知识, 有效弥补了通用大模型在特定学科中的知识断层; 在冶金工艺优化层面, SteelBERT 及 PDGPT 等通过关联工艺参数与显微组织, 为工业级黑盒问题的协同优化提供了高保真度决策支持; 在多智能体协同框架维度, AtomAgents 和 AgentMat 等系统通过层次化架构调度多种专业工具, 驱动了全流程闭环的智能化演进。最后, 针对当前材料大模型面临的数据稀缺性、物理一致性约束及黑盒可解释性等核心挑战, 提出了未来在多模态融合、物理信息引导以及自主实验室方向的发展展望, 旨在为材料基因工程与智能化研发提供理论参考与技术指南。

关键词: 材料科学; 人工智能; 材料大模型; 智能体

中图分类号: TB301 文献标志码: A 文章编号: 0454-5648(2026)07-2363-17

网络出版时间: 2026-04-23



材料科学作为支撑人类社会技术进步与产业变革的基石, 从信息时代的半导体芯片到新能源领域的动力电池, 新材料的发现与应用成为推动科技前沿发展的核心引擎。然而, 传统材料研发长期依赖“试错法”与研究人员的经验直觉, 研发周期长且投入成本高。随着高通量实验技术与高通量计算的普及, 材料研发正经历从“个案研究”向“数据驱动”的范式转变。尽管材料基因组计划(Materials Genome Initiative, MGI)的推进积累了涵盖结构、性能与工艺的各类大规模科学数据库, 但如何从这些复杂、多模态的海量数据中高效挖掘潜在规律, 并将跨学科的信息整合, 实现材料大数据技术的应用, 依然是材料科学领域的严峻挑战。

为应对数据规模的爆发式增长, 早期的材料研发开始引入以传统机器学习(如随机森林、支持向量机等)为代表的“小模型”算法。研究者利用已有数据库对特定材料体系进行训练, 通过构建结构-性能的

映射关系, 显著加速了材料筛选的效率。然而, 这类早期模型往往局限于单一任务或特定材料类别, 且严重依赖人工标注的高质量数据, 在面对跨学科逻辑推理、复杂实验协议生成等高阶需求时显现出泛化性不足的局限。

近年来, 人工智能技术的浪潮, 特别是以大语言模型(Large Language Models, LLMs)为代表的大模型技术^[1-5], 正在重塑材料科学的研究版图。这标志着材料科学从传统的局部数据拟合时代, 迈向人工智能驱动(AI-Driven)的新范式。大模型不再是局限于针对特定指标的判别式工具, 而是通过在海量科学文献、物理化学代码及实验协议基础上的预训练, 实现深度理解复杂材料学机理、自主规划实验路径乃至生成全新材料设计蓝图的目标。这一转变为解决材料科学中长期存在的“高通量制备”与“低通量表征”之间的矛盾, 以及从“目标性能”反推“微观结构”的逆向设计难题, 提供了可靠的解决方案。

收稿日期: 2025-12-30。 修订日期: 2026-01-30。

基金项目: 国家自然科学基金(92463311)。

第一作者: 张林桐(1996—), 女, 硕士, 工程师。

通信作者: 白 洋(1979—), 男, 博士, 研究员。

Received date: 2025-12-30. Revised date: 2026-01-30.

First author: ZHANG Lintong (1996—), female, Master, Engineer.

E-mail: ltZhang@ustb.edu.cn

Correspondent author: BAI Yang (1979—), male, Ph.D., Research Fellow.

E-mail: baiy@mater.ustb.edu.cn

当前,大模型技术在材料科学领域的应用呈现出多维度增长。姜峒伯等^[6]锁定钢铁传统流程工业,从数据提取与保护、性能建模与预测、材料逆向设计及智能冶炼等4个应用场景对大模型技术进行综述与展望;闫姿霓等^[7]侧重于 LLMs 底层逻辑的系统性梳理,深度解析了自然语言处理(Natural Language Processing, NLP)技术演进如何通过提升模型的可解释性来重塑科研人员的物理直觉;王飞跃等^[8]站在人工智能驱动的科学研究的(AI for Science, AI4S)战略高度,从数学、物理、生物、材料等角度简要综述了 AI4S 的发展现状;罗丹宇等^[9]系统剖析了大语言模型在材料领域中知识增强、多模态数据融合及工具协同调用等3个方向的最新进展,阐述了大语言模型推动材料研发流程的技术演进;冯建发等^[10]则从数据基础设施层面强调了 AI 技术对静态材料数据库的激活作用,指出数据库正由传统的存储中心向 AI 创新驱动的“智能中枢”演进。研究人员分别从应用领域、微观表征、AI4S 范式、宏观生产及材料数据库治理等综述视角进行了讨论。

本文侧重在“AI+材料科学”这一交叉领域对大模型进行系统性综述讨论。首先,深入剖析大模型赋能材料科学的理论基础,阐释其如何跨越传统机器学习的限制,实现对复杂材料问题的深度理解。进而,重点论述材料大模型在材料研发整个链路中的核心应用场景,包括晶体结构设计、性能预测与基础知识挖掘、具体材料垂类大模型、冶金工艺优化及多智能体系统(Multi-Agent Systems, MAS)。最后,探讨这一新范式将如何颠覆传统的材料研发逻辑。

1 材料大模型训练框架

通用大模型是基于海量、跨领域且非结构化的

多源数据,通过超大规模自监督预训练,构建出具备广泛通用知识、复杂逻辑推理和多任务泛化能力的深度学习系统。这类模型的核心价值在于能够作为上游底座,支撑下游无数细分领域的应用,而无需为每个任务从零训练。在分类上,通用大模型主要由大语言模型、多模态大模型以及单模态大模型(如纯视觉或纯音频模型)三大支柱构成。其中,以 GPT-4 或 Mistral 为代表的大语言模型专注于人类自然语言的建模与生成,通过对数万亿词元的预测学习人类的逻辑与常识;而 GPT-4o 或 Gemini 等多模态模型则更进一步,将文本、图像、语音乃至视频在统一的语义空间内进行融合映射,实现了跨感官的理解与交互。

材料大模型则是通用大模型技术在物质科学领域的深度垂直化与物理具象化,其定义是利用材料科学领域的专业文献、实验数据、晶体/分子结构以及物理化学定律进行预训练,旨在理解并精通“结构-工艺-性能”这一材料科学核心规律的专业智能系统。与通用大模型学习人类常识不同,材料大模型的核心是学习物质世界的“第一性原理”和微观演化逻辑。

受限于开放域语料的语义范畴,通用大模型在处理材料科学等高专业门槛领域时,常因缺乏领域知识的硬约束而诱发科学幻觉,导致生成的分子形态或晶体拓扑结构偏离基本物理准则。这种功能失效的根本原因是由于通用文本训练范式难以使模型自发解耦并内化材料学中跨尺度的知识维度。

为了构建具备物理保真度的材料大模型(图1),需采用自监督预训练与任务导向型指令微调相结合的技术路径,将材料科学的专业语义特征深度注入模型参数空间,从而实现对复杂科研指令的精准解析与逻辑执行。此外,为确保模型输出在热力学稳定性及结构合法性上与客观物理规律保

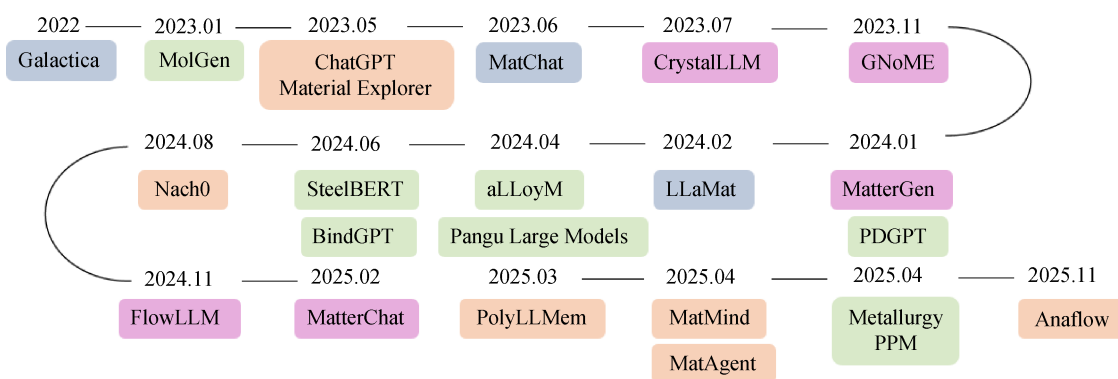


图1 材料大模型汇总

Fig. 1 Summary of material foundation models

持高度一致, 引入强化学习框架^[11]已成为模型逻辑对齐的核心手段。

在材料大模型的演进中, 存在 2 种主流的技术路径: 一是以 CrystaLLM 为代表的语言建模路径, 通过将晶体结构转化为序列化文本, 进而利用 Transformer 架构捕获原子排列的统计规律; 二是以 GNoME(Graph Networks for Materials Exploration)^[12]为代表的几何深度学习路径。尽管 GNoME 并非基于大语言模型的自回归生成机制, 但其基于主动学习(Active Learning)^[13]构建的“预测-计算验证-模型演进”闭环, 在功能逻辑上与强化学习的反馈优化机制异曲同工。该方法通过不断引入高精度第一性原理计算作为外部物理环境的反馈信号, 动态筛选并补充高价值数据以修正模型偏差, 这种物理反馈驱动的迭代优化策略有效地弥补了纯数据驱动模式的局限, 从而确保了所发现的新材料具备极高的热力学稳定性与物理准确度。

本节将从预训练范式、指令微调架构及强化学习等几方面, 系统论述将材料科学先验知识结构化以及映射至大模型参数空间的实现路径, 以期达成模型推理逻辑与材料学底层机理的深度融合。在目前大模型的训练流程中, 指令微调通常通过有监督微调(Supervised Fine-Tuning, SFT)的方法来实现。

1.1 自监督预训练

预训练阶段的本质在于通过非监督表征学习, 从海量非结构化数据中挖掘具备泛化能力的统计特征。在当前的大模型研究中, 这一过程通常遵循基于条件概率预测的自回归架构(如 GPT 系列)或基于上下文重构的掩码语言建模路径(如 BERT^[14-15])。

尽管以 LLaMA、GPT-4 为代表的通用大模型凭借其强大的记忆与模式匹配能力, 能够关联化学键

结、官能团名称等浅层语义符号, 但是这种基于统计概率的文本生成能力并不等同于对物理世界的深度理解。由于缺乏对物理守恒定律的深层逻辑内化, 在处理复杂的晶体结构生成任务时, 模型往往难以自发维持物理一致性, 从而容易产生符合语法规范但违背热力学常识的科学幻觉。因此, 针对材料领域的增量预训练显得较为关键。例如 MatterChat 通过将全分辨率的原子结构信息与预训练大模型进行深度对齐, 显著提升了模型对物质微观结构与宏观性能之间复杂映射逻辑的理解力^[16]。在材料属性预测(如形成能、带隙)和科学推理任务中, 它能够将原子级的几何细节与具体科研情境相结合。在处理复杂的材料合成路径分析时, 它能给出比 GPT-4 更具物理真实性的步骤建议, 展现出深厚的结构感知推理能力; CrystaLLM^[17-18]另辟蹊径, 利用 230 万条晶体学信息文件(Crystallographic Information File, CIF)^[19]进行预训练, 使模型能够理解晶体的空间点群对称性与原子占位规则。这种将 3D 结构映射为序列语言的方法, 使模型在 MP-20 测试集上达到了 58.7% 的单次预测准确率, 具备了自回归生成全新晶体结构的能力。LLaMat^[20]通过在超 120 万篇材料科学文献及数百万个结构化 CIF 上进行大规模持续预训练, 将语言逻辑与晶体几何对称性进行了底层融合。它打破了自然语言与科学编码之间的壁垒, 使得模型不仅能“读懂”文献, 还能“读通”原子排布, 显著提升了模型在处理非结构化科研文本时的物理真实性约束。在逆向设计与晶体结构重构任务中, LLaMat 表现出远胜于通用模型如 Llama-2 的科学幻觉抑制力与结构推演精度。Galactica^[21]通过整合论文、代码、数学公式与化学序列, 构建了宏大的科学知识网络。在处理复杂科学任务时, Galactica

表 1 材料大模型的 3 种训练方法

Table 1 Three training methods for material foundation models

Models	Training Methods	Dataset	Date
Galactica ^[21]	Pre-training	Corpus of 106 billion tokens, including arXiv, Wikipedia, PubChem ^[22] , and UniProt ^[23]	Nov 2022
CrystaLLM ^[17]	Pre-training	OQMD ^[24] , Materials Project ^[25] , COD ^[26] (approx. 1 million CIF files)	Dec 2023
Nach0 ^[27]	Pre-training	Paper abstracts + Chemical molecular SMILES + Crystal SELFIES	Aug 2024
FlowLLM ^[28]	Pre-training	MP-20 (Crystal structure subset of Materials Project)	Nov 2024
MatChat ^[29]	SFT	Materials string (approx. 5000 material instruction entries)	Oct 2023
PolyLLMem ^[30]	SFT	PolyInfo, HOPV, polymer SMILES/SELFIES sequences	Mar 2024
aLLoyM ^[31]	SFT	Alloy-Instruction (approx. 40 000 alloy composition-process instructions)	Apr 2024
PDGPT ^[32]	Pre-training + SFT	Magnesium alloy phase diagram data + related scientific literature	Jan 2024
LLaMat ^[20]	Pre-training + SFT	1.2 million full-text material science papers + CIF structure text	Feb 2024
MatterChat ^[16]	Pre-training + SFT	55 000 full-text materials papers + 4.2 million abstracts + multi-modal images	Feb 2025
MatMind ^[33]	Pre-training + SFT	Cross-scale inorganic non-metallic materials database + experimental process flows	Apr 2025
MolGen ^[34]	Pre-training + RL	Over 100 million unlabeled molecular SELFIES sequence data ^[35-37]	Jan 2023
BindGPT ^[38]	Pre-training + RL	208 million 3D molecular conformations + GEOM-DRUGS dataset ^[39]	Jun 2024

通过在推理过程中插入隐藏的工作内存标记,模拟科研人员在草稿纸上的推导过程,显著提升了模型在物理方程推演和多步数学证明中的逻辑一致性。

1.2 指令微调

纵然基于海量材料学科语料的深度预训练能够促使大模型初步捕获领域表征规律与底层语义特征,但在针对特定下游科研任务的逻辑演化层面,模型依然面临精准优化瓶颈与泛化性约束。如果说预训练是“博览群书”,那么指令微调则是通过任务引导,将模型的通用能力转化为解决特定科研问题的“专家技能”。

在实际科研任务驱动下,大模型需具备对自然语言指令的高度敏感性与执行力,故而引入针对性的 SFT 成为强化其应用适配度与推理精度的核心路径。通过显式构建任务驱动的指令集,模型能够实现从“无约束文本预测”向“目标导向意图解析”的范式转变,从而在面对特定科研指令时,精准激活预训练参数空间内的逻辑链条与格式规范。

此外,依托上下文学习机制,可在不改变模型权重的条件下,通过领域知识提示引导模型实时检索材料学先验图谱,实现对关键理化信息的特征聚焦。当前,大模型在材料科学领域的渗透已覆盖组分设计、拓扑结构演化、物性定量预测及制备工艺全流程优化等维度。针对不同维度的应用需求,指令微调的工程实践侧重于对高专业门槛材料表征语言的深度编解码,以及在训练策略上实现跨模态材料特征与通用语义语料的协同配比与动态平衡。

在材料领域的 SFT 实践中,指令集的构建不再局限于通用的文本交互,而是侧重于多模态材料表征的语义对齐与物理逻辑的显式注入。针对不同材料的特点,研究者设计了差异化的指令微调策略:MatChat^[29]设计了结构化的“晶体结构→电子性质”指令模板,通过引入 CIF 数据的特征描述,实现了从微观几何构型到宏观结构的跨模态预测;对于有机高分子材料,PolyLLMem^[30]针对聚合物长链结构的复杂性,构建了涵盖化学命名、属性预测及合成描述的多任务指令集,使得模型在简化分子线性输入规范序列(Simplified Molecular Input Line Entry Specification, SMILES)或自引用嵌入式字符串序列(Self-Referencing Embedded Strings, SELFIES)与高分子物性之间建立精准的化学语义映射;在冶金工艺优化方面,aLLoyM^[31]引入了思维链机制(Chain of Thought, CoT),引导模型在输出合金相组成之前,先生成包含热力学判据与相律约束的中间推理步

骤,从而有效抑制了模型在处理复杂相变逻辑时的幻觉现象。这种特定领域的指令工程,实现将通用的语言推理能力转化为符合材料科学范式的专业预测技能。

1.3 强化学习

为确保模型输出不仅在语义上连贯,且在物理上切实可行,需要引入强化学习将模型输出与客观规律对齐。不同于通用大模型依赖人类反馈强化学习^[40]来契合人类对话习惯,材料大模型的核心在于构建基于物理反馈的强化学习(Reinforcement Learning from Physical Feedback, RLPF)^[41]。这一框架将大模型作为智能体(Agent),将材料的生成过程建模为马尔可夫决策过程,通过与外部物理环境的交互,强制模型输出遵循热力学稳定性、电荷中性及空间群对称性等硬性约束。其优化目标 $J(\phi)$ 是使生成的材料轨迹在物理评估下的累积回报 R 最大化,即:

$$J(\phi) = E_{\tau \sim \pi_{\phi}(\tau)} [R(\tau)] \quad (1)$$

其中: $J(\phi)$ 是目标函数; $\pi_{\phi}(\tau)$ 是在参数 ϕ 控制下,智能体生成完整轨迹 τ 的概率密度函数; $R(\tau)$ 是智能体获得的总累积回报; $E_{\tau \sim \pi_{\phi}(\tau)} [\bullet]$ 是对所有可能生成的轨迹,按照其出现的概率期望。

因此,首要环节是多维度物理奖励函数 $R(\tau)$ 的精细化设计。由于材料的发现本质上是在势能面上寻找全局或局部极小值的过程,奖励函数必须整合结构合法性、能量稳定性和目标性能三大维度。在结构合法性层面,系统利用化学价键理论和空间几何约束,对生成序列中的原子重叠或配位异常给予极大的负值惩罚;在稳定性层面,模型通过接入快速机器学习势函数或自动化第一性原理计算引擎,将计算得出的形成能(Formation Energy)或凸包能(Energy Above Hull)转化为奖励信号,能量越低则奖励越高。这种客观的反馈机制使得模型能够识别出那些看似语义合理但在热力学上不稳定的“幻觉结构”。

在策略优化阶段,基于近端策略优化(Proximal Policy Optimization, PPO)的对齐技术是维持训练稳定性的核心。由于材料化学空间具有极高的维度和离散性,模型在强化学习早期容易陷入局部最优解(例如反复生成极其简单的稳定晶体)或发生“模式坍塌”。通过引入 KL 散度作为约束项,PPO 算法^[42]能有效限制策略更新的幅度,确保微调后的模型在追求高物理奖励的同时,依然保留预训练阶段习得的大规模材料学常识和跨尺度表征能力。这种约束

机制平衡了模型的探索与利用, 使其在保持生成多样性的前提下, 精准地向高稳定性、高功能性的候选区域收敛。

此外, 针对材料研发中高精度计算和实验验证的高成本痛点, 离线强化学习^[43-44]与主动学习的集成应用构成了大模型自我进化的闭环。通过整合历史数据库中的负样本以及高通量计算产生的不稳定结构, 离线强化学习赋予了模型预判“物理禁区”的能力。当模型部署于逆向设计任务时, 它能自动调用外部模拟软件(如 VASP 或 LAMMPS), 通过小批量的实时物理反馈不断迭代其参数。这种“生成-验证-奖励-进化”的闭环范式, 不仅解决了传统方法中从性能目标到结构实现的逆推难题, 更标志着材料研发从依赖研究人员直觉的“黑箱预测”演进为由物理规律驱动的“智能发现”。

通过上述强化学习技术的深度耦合, 材料大模型超越简单的文本模仿, 实现在受控的物理规律约束下进行创造性探索。这种以物理一致性为导向的训练框架, 使材料大模型在新能源储能、超导材料及半导体研发等前沿领域展现出一定的底层驱动力。

2 应用场景

材料大模型主要针对材料研发流程中的特定节点实现效率跃升。其在材料研发领域的应用从单一研发环节的辅助工具, 逐步拓展到局部研发闭环的

流程优化, 最终向全流程智能化研发体系迈进, 应用场景涵盖晶体结构设计、性能预测与基础知识挖掘、具体材料垂类大模型、冶金工艺优化及多智能体协同框架等多重领域。

2.1 晶体结构设计发现

这一类模型的目标是寻找新材料, 主要解决“如何从无到有生成稳定的新晶体结构”的问题, 其边界在于结构稳定性预测与生成。

GNoME^[12]是由 Google DeepMind 团队于 2023 年底发布的一款颠覆性的材料大模型。它被誉为材料领域的“AlphaGo”^[45], 其核心贡献是专注于发现能够稳定存在的新型无机晶体结构, 将人类已知稳定无机晶体的数量提升了 10 倍以上, 为后续实验合成提供高质量的候选名单。

其运行的核心逻辑建立在一种高效的主动学习闭环之上。GNoME 首先通过两种管道生成海量的候选结构: 一种是对已知稳定晶体进行原子替换的结构管道, 另一种是基于化学成分进行随机排列组合的成分管道。随后, 模型会实时预测这些候选结构的形成能, 并筛选出那些位于凸包之下、具备热力学稳定性的潜在新材料。为了确保预测的严谨性, 这些被筛选出的结构会交由高精度的密度泛函理论(Density Functional Theory, DFT)进行量子力学计算验证。验证后的数据无论是成功还是失败, 都会作为新的训练样本回馈给模型, 使其在不断的自我迭代中提升对稳定结构的识别精度(图 2)。

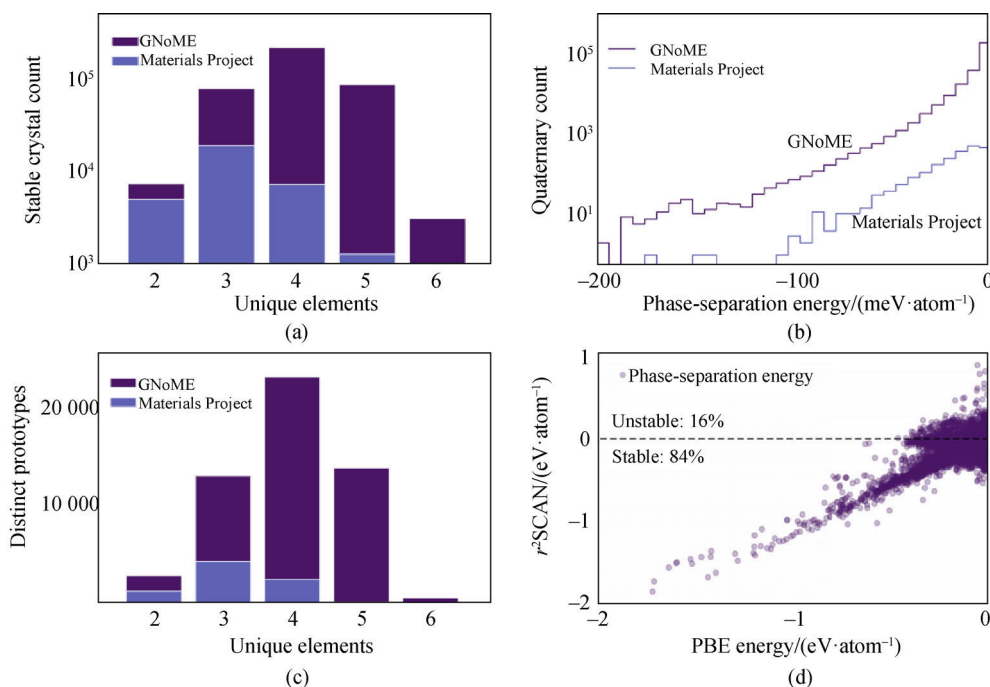


图 2 GNoME 大模型已发现稳定晶体的总结^[12]

Fig. 2 Summary of stable crystals discovered by GNoME^[12]

由美国劳伦斯伯克利国家实验室的 Tang 等^[16]研发的 MatterChat, 是一个能够同时理解材料原子结构和自然语言的大模型(图 3), 它将材料结构数据和文本输入统一到一个内聚模型中, 采用桥接模块,

将预训练的机器学习原子间势与预训练的大模型有效地对齐, 从而降低培训成本并提高灵活性。研究结果表明, MatterChat 显著提高了材料性能预测和人机交互的性能, 超过了 GPT-4 等通用大模型。

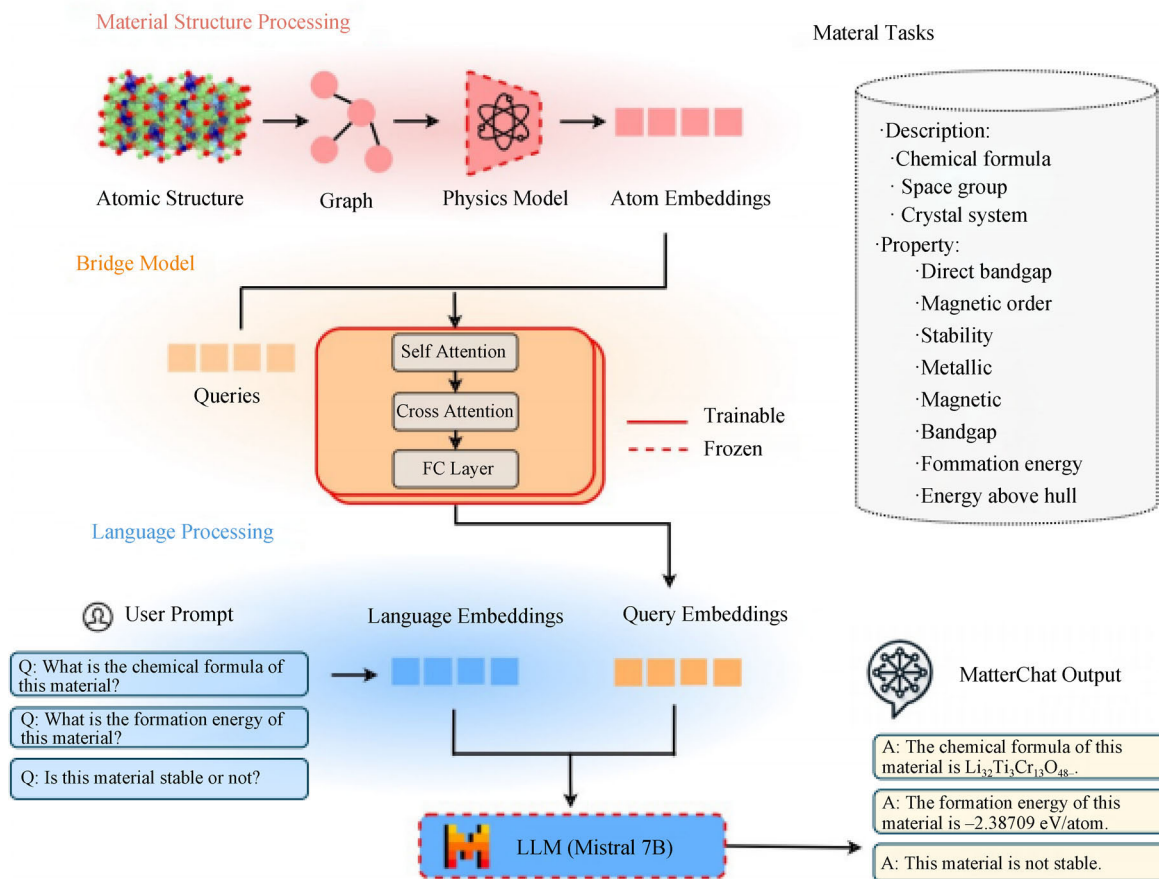


图 3 MatterChat 大模型架构图^[16]

Fig. 3 Architecture diagram of the MatterChat large model^[16]

在晶体材料结构设计方面, 英国雷丁大学研发团队研发的 CrystaLLM 大模型^[17], 将 CIF 进行自回归大型语言建模, 并通过自监督预训练捕获原子排列的统计规律, 从而实现了对于未知无机化合物结构的有效预测。然而, 这种基于字符序列的自回归解码范式虽然在处理离散语义上表现卓越, 但在刻画晶体结构连续的空间几何约束与物理对称性方面仍存在局限。

为了突破这一瓶颈, FlowLLM^[28]引入了黎曼流匹配(Riemannian Flow Matching, RFM)技术^[46], 将生成过程从简单的文本补全演进为在图形表示空间中的连续变量优化。与 CrystaLLM 侧重的因果解码范式不同, FlowLLM 能够将化学计量比与空间群对称性先验深度嵌入概率采样过程中, 实现了从元素组合到平衡构型的端到端映射, 在产生热力学稳定

构型的速率方面较现有基准提升了逾 3 倍以上, 并显著削减了约 50%的后验计算冗余。

2.2 性能预测与基础知识挖掘

这一类模型的目标是打造“科研数字助理”, 作为科研人员的辅助工具, 能够从海量文献中提取知识并进行属性预测。

松山湖材料实验室构建了 MatChat 大模型^[29], 它通过检索增强生成技术, 对接 28 万篇冶金论文知识库(含连铸、轧制等 12 类工艺文献), 基于 LLaMA 2-7B^[47]微调, 采用“问题解构-文献检索-结论合成”三步法, 针对“连铸结晶器冷却水量优化”问题, 通过提示工程精准匹配文献中的“冷却强度-坯壳厚度-裂纹率”关联数据, 答案文献支撑率 100%, 在钢水脱硫工艺参数匹配任务中准确率较 GPT-4 提升 42%, 为材料科学研究者提供精准的知识问答。

Meta AI 团队推出了 Galactica 大模型^[21], 旨在通过构建一种专门针对科学领域的生成式预训练 Transformer 架构, 实现对科学知识的存储、组合与逻辑推理。该数据集包含约 1060 亿个词元, 涵盖了 4800 万篇学术论文、800 万本教科书及数以百万计的科学参考资料。它将晶体结构文件(CIF)、

化学分子式(SMILES)以及数学公式(LaTeX)统一编码为模型可理解的序列(图 4), 这意味着科研人员可以通过自然语言直接与材料的微观结构进行交互, 为材料基因组工程提供了一个强大的语义接口, 标志着材料大模型从通用搜索向知识重组的范式转移。

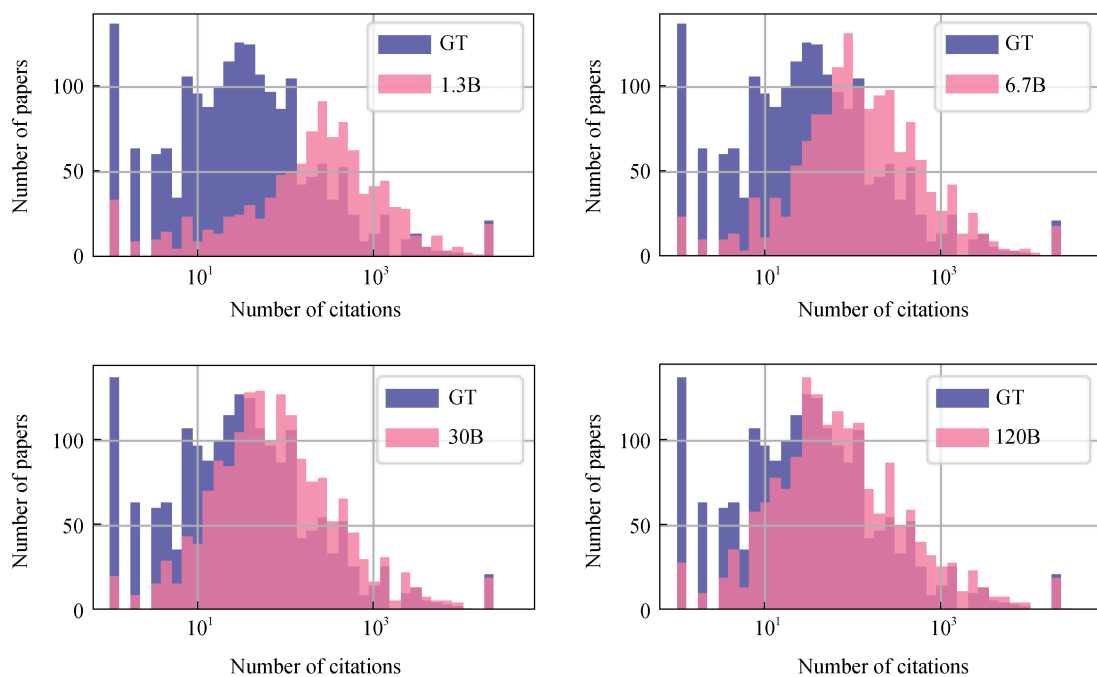


图 4 Galactica 大模型预测论文引用分布结果图^[21]

Fig. 4 Galactica large model prediction results for paper citation distribution^[21]

LLaMat^[20]是由来自卡内基梅隆大学(CMU)、麻省理工学院(MIT)及部分开源社区研究者联合开发的材料大模型(图 5)。LLaMat 采用了 Meta 发布的 Llama-3 架构作为起点, 但在预训练阶段加入了海量的 CIF 和结构化化学数据。它的核心边界在于多模态结构的语义对齐。LLaMat 的开发重点在于让大模型获得“物理常识”, 如理解晶体点阵的周期性与能带结构之间的因果律。作为一个通用底座, LLaMat 的实现效果体现在其极强的可迁移性上, 它被广泛用作其他高阶应用(如材料智能体 MatAgent)的推理大脑, 支持从文本描述直接生成符合物理规则的晶体结构。

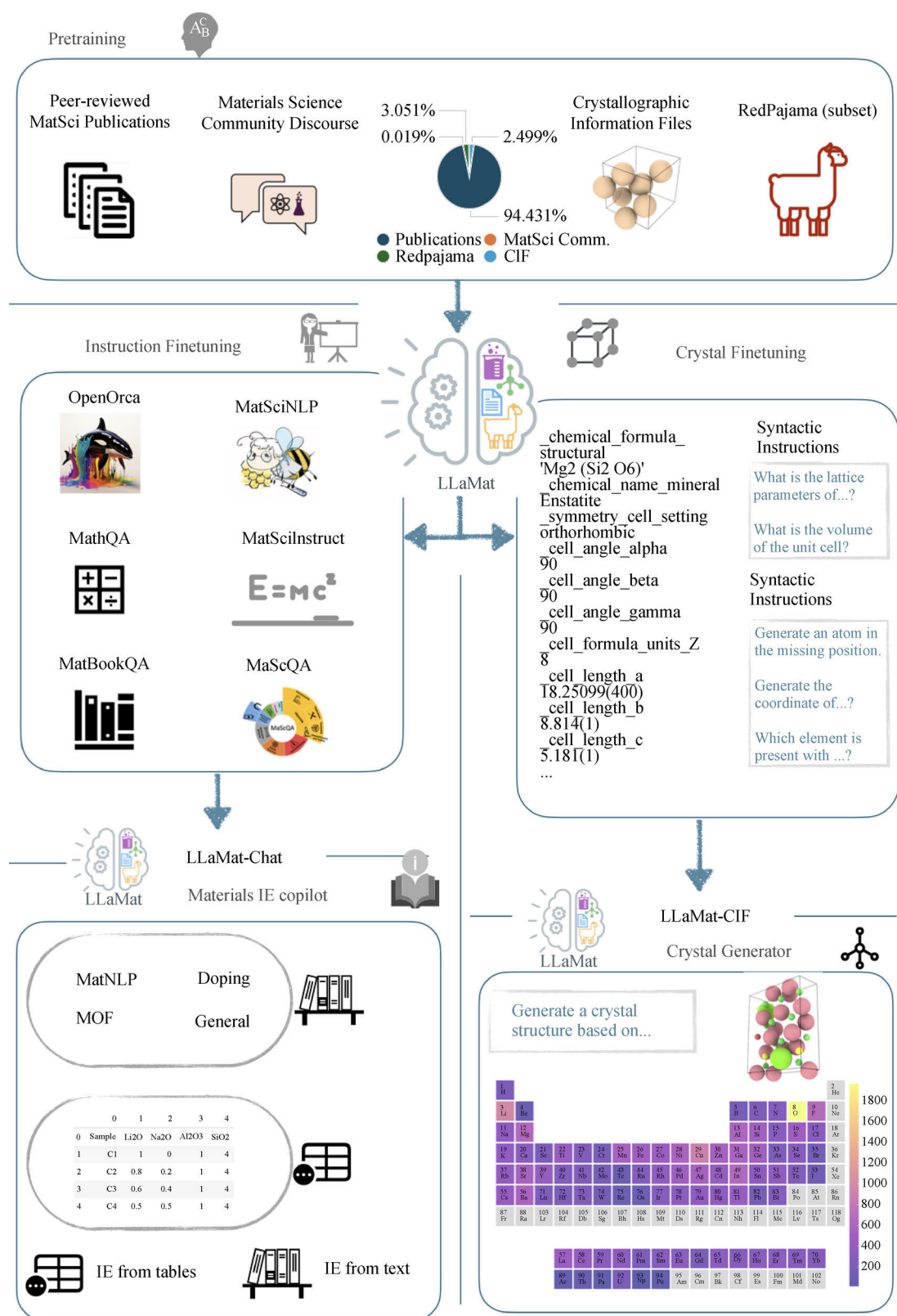
2.3 具体材料垂类大模型

由于不同材料的物理规律差异巨大, 因此需要构建具体材料的垂类大模型, 从而提供更精准的领域参数修正。

在高分子材料研发领域, PolyLLMem^[30]作为专用的垂类大模型, 填补了材料大模型在非晶态、长链有机高分子领域的空白, 使用 2.9 万条数据实现

了超越传统模型的效果。在处理具有高维度搜索空间的带隙预测任务中, PolyLLMem 的决定系数(R^2)稳定在 0.88 以上。在预测玻璃化转变温度指标时, PolyLLMem 的平均绝对误差(Mean Absolute Error, MAE)降低至 12~15 K, 相比于传统模型, 预测精度提升了 20%~35%。

在金属有机框架(Metal organic Framework, MOF)领域, Zheng 等^[48]基于大模型 GPT-4V 的图像解析能力, 实现了数据的挖掘应用。该方法实现了从 346 篇原始文献到 6240 组视觉特征资产的自动化迁移, 利用自然语言处理驱动的多模态感知框架(如 GPT-4V), 对粉末 X 射线衍射(PXRD)、氮气吸附等温线及热重分析(TGA)等非结构化实验图表执行了数字化的逆向工程。此流程不仅能够高保真地提取孔隙率、结晶度等关键理化描述符, 更构建了实验观测值与理论预测间的协同校验机制, 通过精准识别两者间的演化偏差, 为材料体系的逻辑优化提供了数据驱动的启发式指导。

图 5 LLaMat 大模型架构图^[20]Fig. 5 Architecture diagram of the LLaMat large model^[20]

具体在 MOF 的工业应用方面, Kang 等^[49]基于 GPT-4 大模型研发的 ChatMOF 智能体,通过智能体

规划、工具包执行、评估器验证的 3 层架构(图 6),将金属有机框架的设计流程重构为“目标设定-结构

生成-数据检索-性质预测-遗传优化”的迭代循环, 在保持 87.5%生成精度的同时实现了多目标优化,

不仅实现了对 MOF 的物理性质预测, 而且能够对材料的合成路径提出合理化建议。

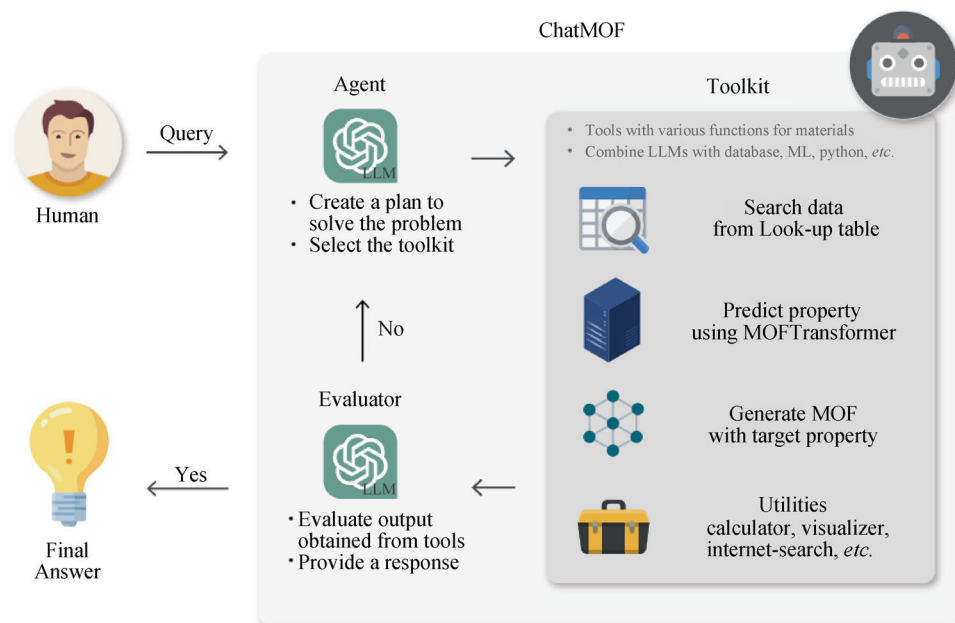


图 6 ChatMOF 智能体架构图^[49]

Fig. 6 Architecture diagram of the ChatMOF agent^[49]

此外, 由德国马普钢铁研究所研究的 ChatGPT Material Explorer(CME)^[50]是一种基于 GPT-4 强大逻辑推理能力构建的材料科学智能体框架。它的出现标志着材料科学进入了“自然语言编程”时代。其核心逻辑在于利用 GPT-4 的 Code Interpreter(代码解释器)或 Function Calling(函数调用)功能, 充当人类科学家与庞大的材料数据库(如 Materials Project、OQMD 等)之间的“翻译官”。传统的材料数据库查询需要复杂的应用程序接口(API)编程知识, 而 CME 允许用户直接使用口语化的指令, 系统会自动将这些指令转化为 Python 代码或 API 请求, 并在后台执行查询、数据清洗与可视化操作。这种模式极大地降低了材料信息学的门槛, 实现了数据的即时挖掘与可视化。CME 不仅能调用数据库, 还能结合 GPT-4 内置的物理与化学常识, 对查询结果进行初步的合理性分析, 过滤掉明显违背热力学常识的异常数据。

在轻合金领域, DeepLight^[51]是一款专门针对铝、镁、钛等轻合金设计的垂类大模型, 通过集成多尺度材料科学知识 with 通用大模型架构, 打破了传统试错在复杂合金设计中的效率瓶颈。该模型通过混合建模技术, 将海量的材料文献数据、高通量 DFT 计算结果与真实的实验表征图像进行多模态融合, 并创新性地引入了物理信息神经网络(Physics-

Informed Neural Networks, PINNs), 将热力学定律与动力学方程嵌入模型底层, 从而确保了预测结果不仅具备统计学意义上的准确性, 更符合材料科学的基本物理逻辑。在实际应用中, DeepLight 建立了从成分配比到加工工艺, 再到微观组织与宏观性能的高维非线性映射, 实现了基于目标性能需求的成分反向设计与热处理工艺精准优化, 不仅显著缩短了新型高强韧轻合金的研发周期, 也为新能源汽车一体化压铸及航天航空构件的轻量化迭代提供了关键的数据驱动支撑。

在无机非金属领域, 中国科学院上海硅酸盐研究所研制的 MatMind 垂类大模型^[33], 标志着材料基因组工程向智能化范式转型。该系统在架构设计上实现了三大关键技术的实质性耦合: 其一, 采用增量预训练驱动的多专家混合机制(Mixture of Experts, MoE); 其二, 构建了集成 CoT^[52]与检索增强生成(Retrieval-Augmented Generation, RAG)^[53]的协同推理逻辑; 其三, 引入了内嵌专家经验的类脑多智能体演化框架。这种多维协同架构有效地弥合了材料研发中微观组织表征与宏观工艺参数间的“尺度断层”, 打破了实验阶段孤立导致的领域知识壁垒。通过对多源异构数据集进行本体化建模与归一化预处理, MatMind 实现了跨尺度数据的一致性映射与逻辑对齐, 在降低计算开销的同时, 显著强

化了模型对复杂物相演变机理的解析精度。该系统具备从实验室尺度探索至工业级规模化生产的全生命周期数据贯通能力, 依托动态更新机制确保了研发全过程的语义衔接。此外, MatMind 通过语义嵌入技术与大规模知识图谱的深度融合, 促进了物理、化学及工程学等跨学科先验知识的非线性集成, 为复杂材料体系的逆向设计与创新研发提供了高保真度的决策支持。

在化学材料领域, Nach0^[27]是由加州理工学院、英伟达与斯坦福大学联合团队研发的一款专为化学

与材料科学设计的垂类大模型(图 7), 其核心创新在于解决了自然语言与化学语言(分子、晶体结构)之间的跨模态语义对齐难题。不同于传统的大模型, Nach0 在底层架构上就实现了“多模态令牌化(Tokenization)”, 它将分子的 SMILES 字符串、晶体的 SELFIES 表示以及描述实验的自然语言文本统一编码为模型可识别的特殊词元。这意味着在 Nach0 的认知世界里, 一段描述药物功能的文字与一个复杂的蛋白质分子结构是平权的, 模型可以流畅地在两者之间进行双向翻译与生成。

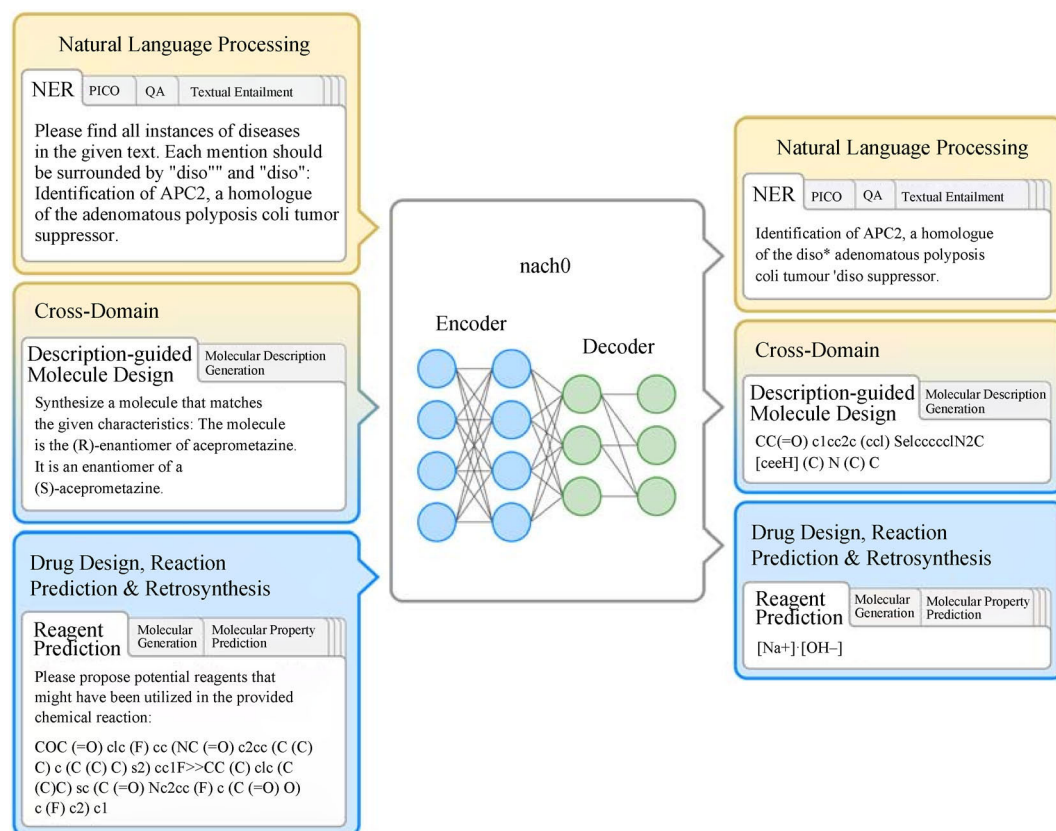


图 7 Nach0 大模型架构示意图^[27]

Fig. 7 Architecture diagram of the Nach0 large model^[27]

在实际应用中, Nach0 展现了惊人的跨任务泛化能力。它不需要针对特定任务进行微调, 即可执行分子属性预测、化学反应产物生成以及晶体结构合成等多种任务。例如, 科研人员可以输入一段关于“高光电转换效率”的文字描述, Nach0 能够直接生成符合该描述的有机光伏分子的化学结构; 反之, 输入一个新发现的晶体结构, 它也能自动生成关于该材料潜在应用场景的文本报告。这种端到端的生成能力, 使其成为生物医药与功能材料领域极具潜力的设计工具。

在半导体领域, 主要体现在缺陷预测与材料设

计上。具体在半导体缺陷分类与预测方面, 格创东智科技公司构建了一个“知识驱动+数据驱动”的双引擎架构 AI FDC^[54], 让工业大模型先去学习设备手册、专家经验等非结构化知识, 智能推荐需要监控的关键参数, 替代了工程师手动筛选的繁重工作。系统再运用轨迹模式识别和多变量分析等算法, 实现对设备异常的精准捕捉和根因分析, 将故障检测从依赖经验的“静态手艺”升级为可自适应、自进化的智能系统。

在芯片设计端, 大模型正在帮助工程师应对日益复杂的电路设计挑战。NVIDIA 推出的 ChipNeMo^[55]

大模型, 通过生成式 AI 技术解决芯片设计过程中极高的知识壁垒与繁琐的工作流问题。研究团队以 Llama-2 为基座模型, 利用超过 240 亿个词元的芯片设计专用数据进行持续训练, 这些数据涵盖了大量的硬件描述语言(如 Verilog)、复杂的内部设计规范、详细的架构文档以及海量的 EDA 工具手册。为了让模型更懂硬件设计, ChipNeMo 定制了领域适配的分词器, 使得分词效率显著提升。这种深度的底层改造, 使其能够深入理解电路逻辑的时序性与并发性。

此外, 具体在设计与自动化方面, 比利时鲁汶大学研发的 Anaflow^[56]智能体创新地使用了多个专门的 AI 智能体(图 8), 并通过迭代推理与反馈闭环持续修正电路参数。研究数据表明, 与传统的遗传算法相比, AnaFlow 在处理复杂信号电路设计时, 其参数收敛速度提升了 30%~50%, 并在特定拓扑下将品质因数提升了约 15%。更重要的是, 整个过程能够为每一次参数调整提供基于电路理论的设计决策依据, 使得设计过程更加透明和高效。

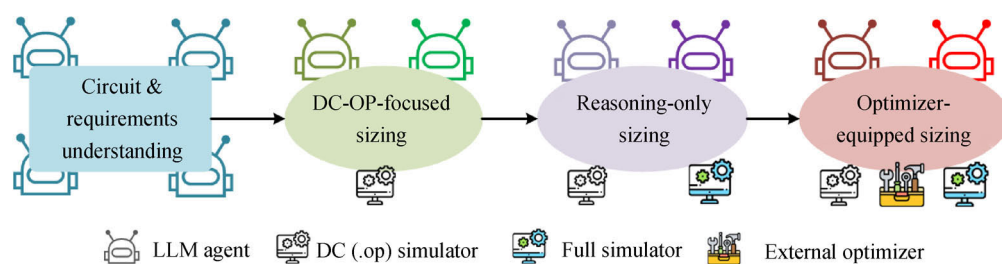


图 8 Anaflow 智能体 4 个阶段概览图^[53]

Fig. 8 Overview diagram of the four stages of the Anaflow agent^[53]

2.4 冶金工艺优化

这一类模型的赛道从实验室到工厂, 其边界在于工艺参数与显微组织、宏观性能的关联, 重点解决冶金过程中的“黑盒”问题。

在冶金工艺优化领域, 智能感知大模型——中国钢研冶金流程感知大模型^[57-58], 被视为一个行业级的智能基座。它不是针对单一参数, 而是通过负责识别数据特征的“感侧大模型”和基于工艺机理进行推理判断的“知侧大模型”的双塔结构, 为整个冶金流程提供多尺度的感知能力。它重点解决了数据生成难题, 通过引入冶金机理, 规模化生成了用于训练的高价值数据, 特别是在解决金相分析、缺陷检测等长尾问题上表现出色。中国钢研已以此为基础, 打造一个完整的行业大模型矩阵。

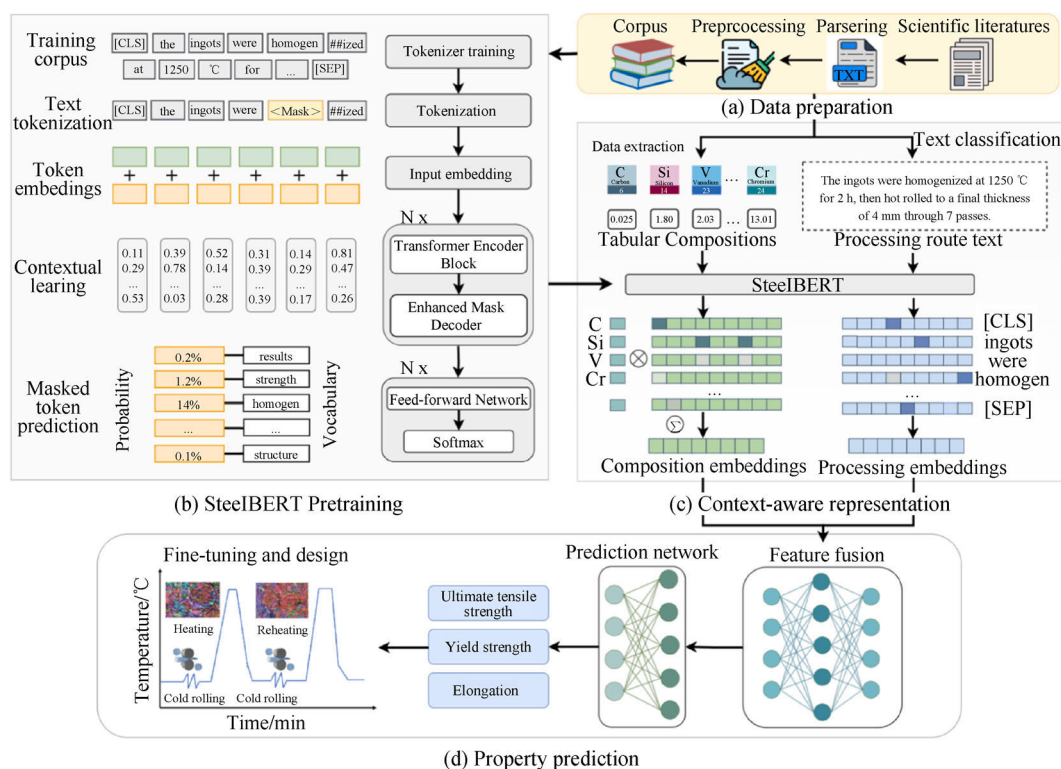
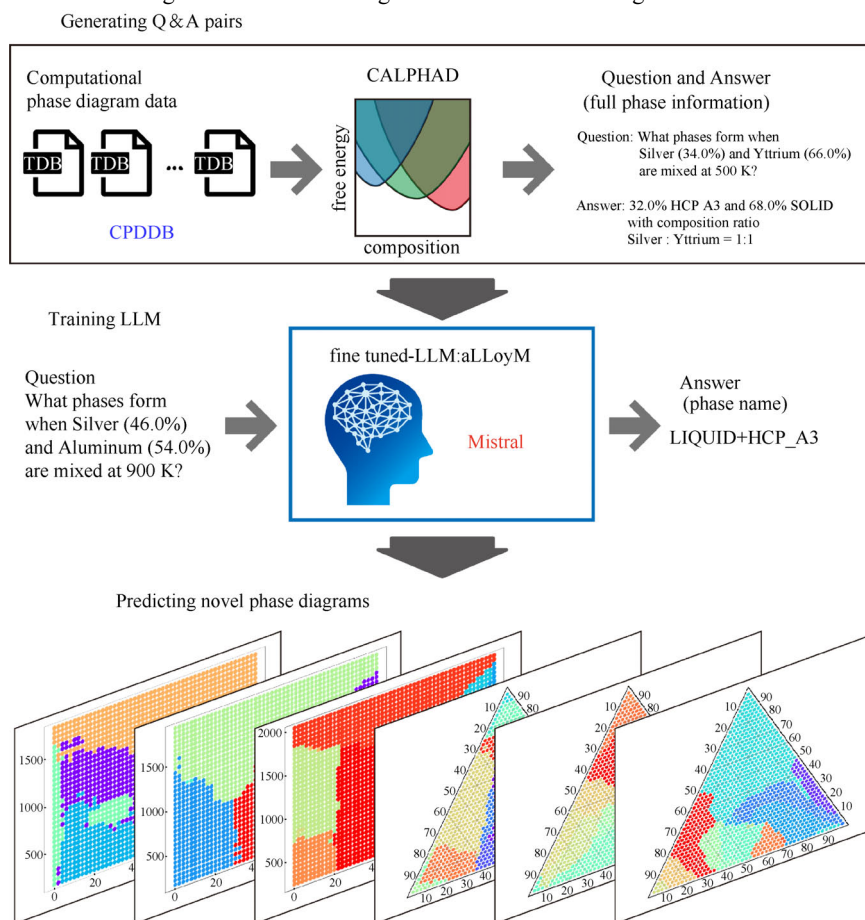
北京科技大学新材料大数据中心与阿里云联合发布的钢铁材料设计大模型 SteelBERT^[59], 基于 DeBERTa^[60]进行预训练(图 9), 并使用了约 420 万篇材料摘要和 55 000 篇钢铁文章全文作为语料库。通过训练, SteelBERT 能够学习到钢铁材料的成分、工艺和性能之间的关系, 从而实现对钢结构性能的准确预测, 在材料研发场景中, 性能预测精度最高可达 90%, 进一步融合阿里通义千问, 打造 SteelScientist 多智能体平台, 实现了从数据挖掘、“成

分-工艺-性能”一体化设计到专业知识智能问答的全链路突破。

由 Oikawa 等^[31]提出的 aLLoyM 大模型 (图 10), 基于 Mistral LLM^[61]微调, 整合 CPD 数据库^[62]与 CALPHAD^[63]评估结果构建问答数据集, 采用“成分编码-相区映射-动态渲染”技术路径, 通过提示工程实现相图数据的自然语言交互解读。在 Al-Zn-Mg、Cu-Ni-Si 等 12 类合金相组成判断中准确率达 92.7%, 较通用模型提升 67%, 并支持生成带相变点标注的动态相图。

由 Yan 等^[32]研发的 PDGPT 大模型, 基于检索增强生成技术构建大模型驱动的相图分析系统, 融合 10 万+计算相图数据, 通过“自然语言提问-文献数据检索-结构化输出”闭环, 用户仅需输入合金成分即可获得相变温度、相组成等关键信息, 即可输出相变温度(MAE≤20 K)及含量占比, 交互式界面使非 CALPHAD 专家的相图分析效率提升 3 倍以上, 其交互式问答设计显著降低了相图分析的专业门槛, 目前已应用于航空铝合金选材优化。

湘钢和华为公司联合提出的盘古大模型^[64-65]展现了点对点解决具体工艺难题的能力。它通过在精炼这一关键环节深度应用, 动态预测钢水成分和温度变化, 直接将预测误差减半, 稳定了产品质量, 创造了显著的经济效益。

图9 SteelBERT 大模型架构示意图^[58]Fig. 9 Architecture diagram of the SteelBERT large model^[58]图10 aLLoyM 大模型架构示意图^[31]Fig. 10 Architecture diagram of the aLLoyM large model^[31]

2.5 多智能体协同框架

在材料研发迈向智能化范式转变的过程中, MAS 正逐渐从单一的辅助工具演化为驱动“设计-模拟-实验-优化”全流程闭环的核心引擎^[66-67]。这一转变标志着材料科学研究进入了由自然语言驱动的自动化协同时代。通过将复杂的大规模研发任务解构为多个可执行的子任务, 并分配给具备特定功能的智能体, 研究人员能够实现跨尺度的知识整合与跨工具链的逻辑调用。在这一框架下, 大语言模型不再仅仅承担文本生成职责, 而是作为“中心大脑”进行任务规划与工具调度。

AtomAgents^[68]是由 MIT 的 Markus Buehler 教授团队开发的一种高度集成的多智能体 AI 框架(图 11)。其核心价值在于通过将大模型的自然语言理解力与物理信息启发式算法深度耦合, 实现了

材料研发从原子设计到循环优化的全流程协同。在设计阶段, 设计 Agent 将用户的模糊需求解构为物理约束, 协调具备材料学知识的智能体提出具有理论支撑的原子组分建议; 随后, 模拟 Agent 自主编写并运行分子动力学或量子力学代码, 在虚拟空间中对设计出的结构进行极限测试; 进入实验环节后, 数据 Agent 能够同化处理实验表征图像与传感器数据, 通过对比模拟预测与真实观测的偏差, 识别出微观机制对宏观性能的影响规律; 最后, 系统利用优化 Agent 在多维参数空间内进行启发式搜索, 将识别出的关键物理特征反馈至下一轮设计循环。这种基于多智能体群体智能的闭环机制, 不仅打破了跨尺度模拟的技术壁垒, 更让材料研发从传统试错进化为由物理法则驱动的自主发现范式。

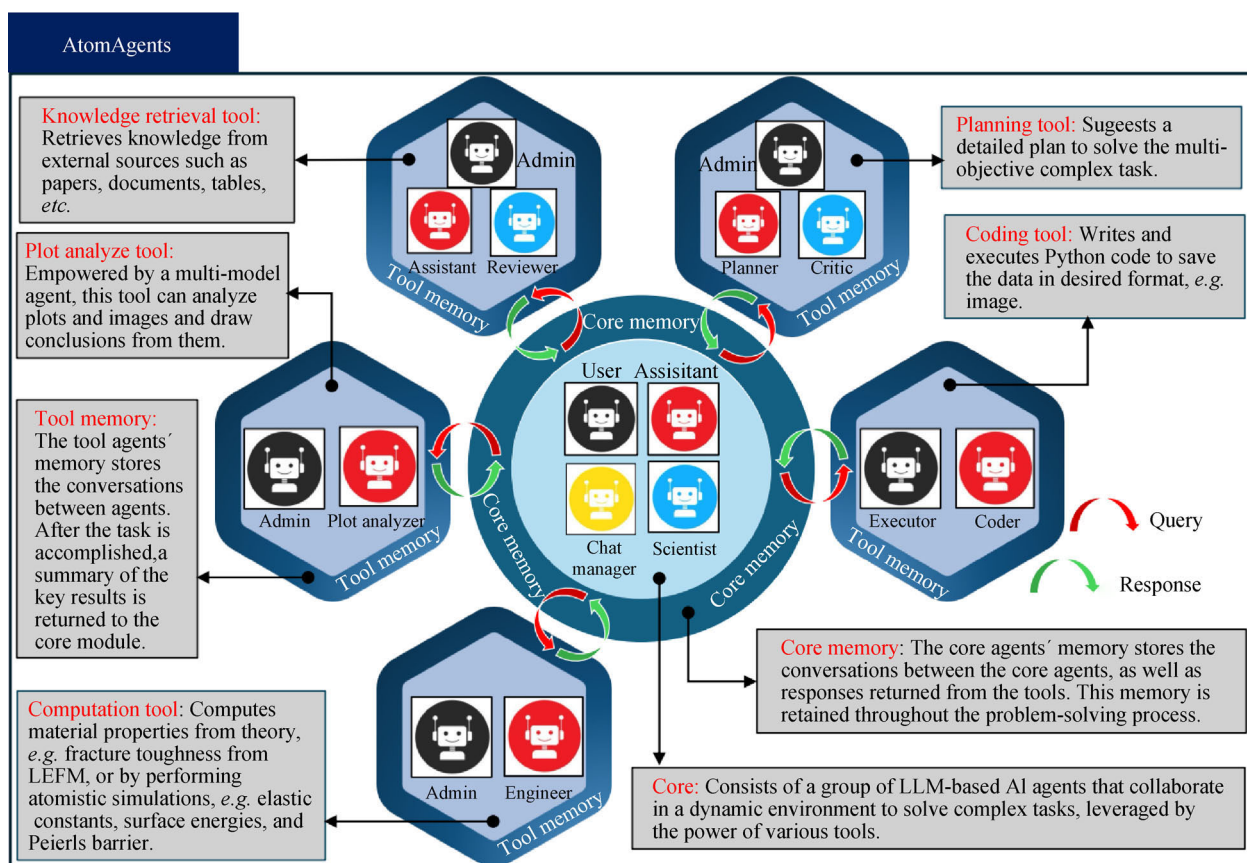


图 11 AtomAgents 多智能体协同框架^[68]

Fig. 11 Collaboration structure of AtomAgents MAS^[68]

由上海交通大学联合小米集团发布的 AgentMat^[69]多智能体, 作为一个专为轻合金设计的自动化研发平台, 其核心在于构建了一个由材料领域的垂类大模型 DeepLight^[51]驱动的多智能体协同框架。在初始的设计阶段, 设计 Agent 能够深入理

解工业级性能指标, 利用其内置的轻合金专业知识库提出候选的化学组分与微观组织方案; 紧接着, 模拟 Agent 会自主调用物理计算工具, 生成相应的输入参数并驱动高保真模拟模型进行性能推演, 从而在原子或介观尺度验证设计的可行性; 当流程进

入实验对接环节, 数据 Agent 会通过标准化的接口整合实验表征结果, 利用其强大的跨模态理解能力, 将实验观测到的宏观缺陷或组织演变与早期的模拟预测进行横向比对; 最终, 优化 Agent 根据实验与模拟之间的偏差进行因果推理, 动态调整搜索空间并引导系统进入下一轮迭代。这种多智能体之间的实时通信与知识共享机制, 彻底改变了传统材料研发中各环节相互孤立的状态, 实现了从市场需求到材料落地的高度自动化集成。

3 结论与展望

现如今, 材料大模型凭借其在海量科学文献与结构化数据中涌现出的深度语义理解与逻辑推理能力, 正以飞快的速度重构材料研发的全生命周期。从自监督预训练到 RLDPF, 材料大模型已不再仅仅是辅助文本生成的工具, 而是进化为能够理解晶体几何约束、遵循热力学规律并自主规划实验路径的科研智能体。

然而, 在材料大模型研发的进程中, 依然存在亟待突破的瓶颈。当前的系统往往面临物理一致性约束不足、训练数据稀缺以及深度学习模型固有的“黑盒”可解释性难题, 导致模型在处理原子尺度演化与宏观性能关联时, 仍难以揭示深层的因果机制。同时, 离散的研发环节与僵化的智能体协作模式, 制约了从“目标设计”到“实验验证”闭环的响应效率, 使得大模型在面对非预期实验现象时缺乏足够的自适应能力。

其中, 针对数据稀缺性问题, 当前范式正从通用建模转向知识迁移。例如, SteelBERT 首先在数以百万计的语料库上进行预训练, 习得原子排列的基本规律; 随后通过迁移学习, 在样本量仅有几百个的特殊钢实验数据集(如高熵合金或特定热处理工艺数据)上进行微调。实验证明, 这种“大规模数据预训练+小规模实验数据微调”的策略, 使模型在实验数据极度匮乏的条件下, 预测精度比传统随机森林等算法提升 30% 以上, 有效解决了从实验室研发到工业应用的数据鸿沟。此外, MatMind 采用的数据增强技术, 对晶体坐标进行微小扰动, 从而增加样本多样性, 显著提升了模型在处理复杂相变预测时的鲁棒性。针对可解释性问题, 通过集成注意力权重可视化技术与物理规则显式编码, 有效地将模型内部的复杂数学逻辑转化为直观的物理图景, 从而在微观原子特征与宏观材料性能预测之间建立桥梁。例如, MatterChat 首先注入原子轨道能级与

电负性梯度等物理描述符, 引导大模型学习具有物理意义的特征空间。通过提取在预测带隙(Bandgap)时的注意力图显示, 使大模型在进行决策时, 自动将高权重分配给活性位点附近的原子轨道特征, 从而与传统化学中的配位场理论高度吻合。CrystaLLM 通过提取多头注意力矩阵中的权重分布, 并将其逆向映射回晶格的 3D 坐标空间, 可以捕捉模型在原子堆垛的能量分布。当系统发生对称性破缺, 可视化结果会显示模型在这些发生位移的原子轨道或化学键上的权重分配, 从而实现对未知无机化合物结构的有效预测。

展望未来, 材料大模型的发展将聚焦于多模态深度融合与物理信息引导的协同进化。通过构建具有因果推理能力的智能引擎, 将微观物理规律与宏观唯象机理融入模型架构, 将是消除幻觉并实现科学发现可解释性的关键。未来的研发体系将趋向于构建由多智能体动态协作驱动的自主实验室, 通过集成实时感知与决策模块, 实现跨尺度、跨学科的知识重组与实验自动迭代。这不仅将从根本上颠覆传统的材料设计逻辑, 更将为材料基因组工程注入强大的智能动力, 引领材料科学进入全流程智能化研发的新时代。

参考文献:

- [1] BROWN T B, MANN B, RYDER N, et al. Language models are few-shot learners[C]//Proceedings of the 34th International Conference on Neural Information Processing Systems. ACM, 2020: 1877–1901.
- [2] VASWANI A, SHAZEER N, PARMAR N, et al. Attention is all you need[C]//Proceedings of the 31st International Conference on Neural Information Processing System, 2017: 6000–6010.
- [3] TOUVRON H, LAVRIL T, IZACARD G, et al. LLaMA: Open and efficient foundation language models[PP/OL]. arXiv (2023–02–27) [2026–04–08]. <https://arxiv.org/abs/2302.13971>.
- [4] 沈洋. 人工智能赋能材料科学[J]. 硅酸盐学报, 2026, 54(1): 1–2. SHEN Yang. J Chin Ceram Soc, 2026, 54(1): 1–2.
- [5] 吴克登, 陈炎, 邓晓彬, 等. AI 驱动的材料发现: 生成式模型与晶体逆向设计进展[J]. 硅酸盐学报, 2026, 54(1): 90–105. WU Kedeng, CHEN Yan, DENG Xiaobin, et al. J Chin Ceram Soc, 2026, 54(1): 90–105.
- [6] 姜峒伯, 朱德鑫, 吴宏辉, 等. 大语言模型赋能钢铁行业: 技术与应用展望[J]. 冶金自动化, 2025, 49(4): 17–35. JIANG Tongbo, ZHU Dexin, WU Honghui, et al. Metall Ind Autom, 2025, 49(4): 17–35.
- [7] 闫姿霓, 饶梓元, 曾小勤. 大语言模型在材料科学领域的研究进展[J]. 中国材料进展, 2025, 44(5): 409–423. YAN Zini, RAO Ziyuan, ZENG Xiaoqin. Mater China, 2025, 44(5): 409–423.
- [8] 王飞跃, 缪青海, 张军平, 等. 探讨 AI for Science 的影响与意义: 现状与展望[J]. 智能科学与技术学报, 2023, 5(1): 1–6. WANG Feiyue, MIAO Qinghai, ZHANG Jumping, et al. Chin J Intell Sci Technol, 2023, 5(1): 1–6.

- [9] 罗丹宇, 温李阳, 陈博, 等. 材料科学大语言模型关键技术及其应用[J]. 中国科学: 信息科学, 2025, 55(7): 1599–1621.
LUO Danyu, WEN Liyang, CHEN Bo, et al. Sci Sin Informationis, 2025, 55(7): 1599–1621.
- [10] 冯建发, 王畅畅, 苏航, 等. 材料数据库的现状与未来: AI 技术引领的创新应用前景[J]. 中国材料进展, 2026, 45(2): 89–101.
FENG Jianfa, WANG Changchang, SU Hang, et al. Mater China, 2026, 45(2): 89–101.
- [11] SUTTON R S, BARTO A G. Finite Markov Decision Processes [M]// SUTTON R S, BARTO A G. Reinforcement Learning: An Introduction. 2nd ed. Cambridge: MIT Press, 2018: 47–71.
- [12] MERCHANT A, BATZNER S, SCHOENHOLZ S S, et al. Scaling deep learning for materials discovery[J]. Nature, 2023, 624(7990): 80–85.
- [13] SETTLES B. Active learning literature survey[R]. Computer Sciences Technical Report 164, University of Wisconsin-Madison, 2009.
- [14] HE K, ZHANG X, REN S, et al. Deep residual learning for image recognition[C]//IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, 2016: 770–778.
- [15] DEVLIN J, CHANG M W, LEE K, et al. BERT: Pre-training of deep bidirectional transformers for language understanding[PP/OL]. V2, arXiv (2019–05–24) [2026–04–08]. <https://doi.org/10.48550/arXiv.1810.04805>.
- [16] TANG Y, XU W, CAO J, et al. MatterChat: A Multi-modal Large Language Model for Materials Science [PP/OL]. V1, arXiv (2024–10–21) [2026–04–08]. <https://arxiv.org/abs/2410.12345>.
- [17] ANTUNES L M, BUTLER K T, GRAU-CRESPO R. Crystal structure generation with autoregressive large language modeling[J]. Nat Commun, 2024, 15: 10570.
- [18] ZENI C, PINSLER R, ZÜGNER D, et al. A generative model for inorganic materials design[J]. Nature, 2025, 639: 624–632.
- [19] BJÖRKMAN T. CIF2Cell: Generating geometries for electronic structure programs[J]. Comput Phys Commun, 2011, 182(5): 1183–1186.
- [20] MISHRA V, SINGH S, AHLAWAT D, et al. Foundational large language models for materials research [PP/OL]. arXiv (2025–01–28) [2026–04–13]. <https://arxiv.org/abs/2412.09560>.
- [21] TAYLOR R, KARDAS M, CUCURULL G, et al. Galactica: A large language model for science[PP/OL]. arXiv (2022–11–16) [2026–04–08]. <https://arxiv.org/abs/2211.09085>.
- [22] KIM S, CHEN J, CHENG T J, et al. PubChem 2023 update[J]. Nucleic Acids Res, 2023, 51(D1): D1373–D1380.
- [23] CONSORTIUM T U. UniProt: The universal protein knowledgebase[J]. Nucleic Acids Res, 2017, 45(D1): D158–D169.
- [24] KIRKLIN S, SAAL J E, MEREDIG B, et al. The Open Quantum Materials Database (OQMD): Assessing the accuracy of DFT formation energies[J]. npj Comput Mater, 2015, 1: 15010.
- [25] JAIN A, ONG S P, HAUTIER G, et al. Commentary: The Materials Project: A materials genome approach to accelerating materials innovation[J]. APL Mater, 2013, 1: 011002.
- [26] DRAXL C, SCHEFFLER M. The NOMAD laboratory: From data sharing to artificial intelligence[J]. J Phys Mater, 2019, 2(3): 036001.
- [27] LIVNE M, MIFTAHUTDINOV Z, TUTUBALINA E, et al. nach0: Multimodal natural and chemical languages foundation model[J]. Chem Sci, 2024, 15(22): 8380–8389.
- [28] SRIRAM A, MILLER B K, CHEN R T Q, et al. FlowLLM: Flow matching for material generation with large language models as base distributions[PP/OL]. arXiv (2024–10–30) [2026–04–08]. <https://arxiv.org/abs/2410.23405>.
- [29] CHEN, XIE, WAN M, et al. MatChat: A large language model and application service platform for materials science[J]. Chin Phys B, 2023, 32(11): 118104.
- [30] ZHANG T R, YANG D B. Multimodal machine learning with large language embedding model for polymer property prediction[J]. Chem Mater, 2025, 37(18): 7002–7013.
- [31] OIKAWA Y, DEFFRENNES G, ABE T, et al. aLLoyM: A large language model for alloy phase diagram prediction[PP/OL]. arXiv (2025–08–10) [2026–04–08]. <https://arxiv.org/abs/2507.22558>.
- [32] YAN Z N, LIANG H Y, WANG J Y, et al. PDGPT: A large language model for acquiring phase diagram information in magnesium alloys[J]. Mater Genome Eng Adv, 2024, 2(4): e77.
- [33] 中国科学院上海硅酸盐所. 人工智能为材料研发提供精准支持[N]. 上海科技报, 2025–04–09(1).
- [34] FANG Y, ZHANG N Y, CHEN Z, et al. Domain-agnostic molecular generation with chemical feedback[PP/OL]. arXiv (2024–03–04) [2026–04–08]. <https://arxiv.org/abs/2301.11259>.
- [35] POLYKOVSKIY D, ZHEBRAK A, SANCHEZ-LENGELING B, et al. Molecular sets (MOSES): A benchmarking platform for molecular generation models[J]. Front Pharmacol, 2020, 11: 565644.
- [36] STERLING T, IRWIN J J. ZINC 15–ligand discovery for everyone[J]. J Chem Inf Model, 2015, 55(11): 2324–2337.
- [37] ZHAO H, YANG Y, WANG S Q, et al. NPASS database update 2023: Quantitative natural product activity and species source database for biomedical research[J]. Nucleic Acids Res, 2023, 51(D1): D621–D628.
- [38] ZHOLUS A, KUZNETSOV M, SCHUTSKI R, et al. BindGPT: A scalable framework for 3D molecular design via language modeling and reinforcement learning[PP/OL]. arXiv (2024–01–06) [2026–04–08]. <https://arxiv.org/abs/2406.03686>.
- [39] FRANCOEUR P G, MASUDA T, SUNSERI J, et al. Three-dimensional convolutional neural networks and a cross-docked data set for structure-based drug design[J]. J Chem Inf Model, 2020, 60(9): 4200–4215.
- [40] Long Ouyang, Jeff Wu, Xu Jiang, et al. Training language models to follow instructions with human feedback [PP/OL]. arXiv (2022–03–04) [2026–04–08]. <https://arxiv.org/abs/2203.02155>.
- [41] WANG J, WANG Z, WANG Y, et al. RL from physical feedback: Aligning large motion models with humanoid control [C]//International Conference on Learning Representations (ICLR), Vienna, Austria, 2026: 1205–1218.
- [42] SCHULMAN J, WOLSKI F, DHARIWAL P, et al. Proximal policy optimization algorithms[PP/OL]. arXiv (2017–08–28) [2026–04–08]. <https://arxiv.org/abs/1707.06347>.
- [43] LEVINE S, KUMAR A, TUCKER G, et al. Offline reinforcement learning: Tutorial, review, and perspectives on open problems.[PP/OL]. arXiv (2020–11–01) [2026–04–08]. <https://doi.org/10.48550/arXiv.2005.01643>.
- [44] FUJIMOTO S, MEGER D, PRECUP D. Off-policy deep reinforcement learning without exploration [C]//36th International Conference on Machine Learning (ICML), Long Beach, USA, 2019: 2052–2062.
- [45] SILVER D, HUANG A, MADDISON C J, et al. Mastering the game of Go with deep neural networks and tree search[J]. Nature, 2016, 529(7587): 484–489.
- [46] LIU X C, GONG C Y, LIU Q. Flow straight and fast: Learning to generate and transfer data with rectified flow[PP/OL]. arXiv (2022–09–07) [2026–04–08]. <https://arxiv.org/abs/2209.03003>.
- [47] TOUVRON H, MARTIN L, STONE K, et al. Llama 2: Open foundation and fine-tuned chat models[PP/OL]. arXiv (2023–07–19) [2026–04–08]. <https://arxiv.org/abs/2307.09288>.
- [48] ZHENG Z L, HE Z G, KHATTAB O, et al. Image and data mining in reticular chemistry powered by GPT-4V[J]. Digit Discov, 2024, 3(3): 491–501.
- [49] KANG Y, KIM J. ChatMOF: An artificial intelligence system for predicting and generating metal-organic frameworks using large

- language models[J]. Nat Commun, 2024, 15: 4705.
- [50] CHOUDHARY K. ChatGPT material explorer: Design and implementation of a custom GPT assistant for materials science applications[J]. Integr Mater Manuf Innov, 2025, 14(3): 276–283.
- [51] DENG W, PAN J W, ZHOU T, et al. DeepLight: Deep lightweight feature interactions for accelerating CTR predictions in ad serving[C]//Proceedings of the 14th ACM International Conference on Web Search and Data Mining. Virtual Event Israel. ACM, 2021: 922–930.
- [52] BOSMA M, CHI E, ICHTER B, et al. Chain-of-thought prompting elicits reasoning in large language models[C]//Advances in Neural Information Processing Systems 35. New Orleans, Louisiana, USA. Neural Information Processing Systems Foundation, Inc. (NeurIPS), 2022: 24824–24837.
- [53] LEWIS P, PEREZ E, PIKTUS A, et al. Retrieval-augmented generation for knowledge-intensive NLP tasks [C]//Advances in Neural Information Processing Systems. Red Hook, NY: Curran Associates, Inc., 2020, 33: 9459–9474.
- [54] 格创东智. 格创东智: AI 驱动制造业向更智能迈进[J]. 现代制造, 2024(12): 35.
- Ge Chuang Dong Zhi. Maschinen Markt, 2024(12): 35.
- [55] LIU M J, ENE T D, KIRBY R, et al. ChipNeMo: Domain-adapted LLMs for chip design[PP/OL]. arXiv (2024–04–04) [2026–04–08]. <https://arxiv.org/abs/2311.00176>.
- [56] AHMADZADEH M, CHEN K, GIELEN G. AnaFlow: Agentic LLM-based workflow for reasoning-driven explainable and sample-efficient analog circuit sizing[C]//2025 IEEE/ACM International Conference on Computer-Aided Design (ICCAD), Munich, Germany, 2025: 1–7.
- [57] LIU Q, CHEN W L, YAKUBOV V, et al. Interpretable machine learning approach for exploring process-structure-property relationships in metal additive manufacturing[J]. Addit Manuf, 2024, 85: 104187.
- [58] LIU C, TANG L X, ZHAO C C. A novel dynamic operation optimization method based on multiobjective deep reinforcement learning for steelmaking process[J]. IEEE Trans Neural Netw Learn Syst, 2024, 35(3): 3325–3339.
- [59] TIAN S H, JIANG X, WANG W R, et al. Steel design based on a large language model[J]. Acta Mater, 2025, 285: 120663.
- [60] P. He, X. Liu, J. Gao, et al. DeBERTa: Decoding-enhanced BERT with Disentangled Attention [C]//International Conference on Learning Representations (ICLR), Vienna (Virtual), 2021.
- [61] JIANG A Q, SABLAYROLLES A, Mensch A, et al. Mistral 7B [PP/OL]. arXiv (2023–10–10) [2026–04–08]. <https://doi.org/10.48550/arXiv.2310.06825>.
- [62] DUNN A, WANG Q, GANOSE A, et al. Benchmarking materials property prediction methods: The Matbench test set and Automatminer reference algorithm[J]. npj Comput Mater, 2020, 6(1): 138.
- [63] SAUNDERS N, MIODOWNIK P. CALPHAD : calculation of phase diagrams : A comprehensive guide[M]. Elsevier, 1998.
- [64] 华为盘古大模型上线: 多行业应用带来显著效益[EB/OL]. (2024–07–08) [2025–06–22]. <https://ai.zol.com.cn/879/8790291.html>.
- [65] 山东能源集团携手华为发布盘古矿山大模型共建智能矿山[EB/OL]. (2023–07–12) [2025–05–10]. <http://www.xinhuanet.com/tech/20230720/97d8b97887584cbb9cd79ab23624898d/c.html>.
- [66] 张颖伟, 高鸿瑞, 张鼎森, 等. 基于多智能体的数字孪生及其在工业中应用的综述[J]. 控制与决策, 2023, 38(8): 2168–2182.
- ZHANG Yingwei, GAO Hongrui, ZHANG Dingsen, et al. Control Decis, 2023, 38(8): 2168–2182.
- [67] 黄昌勤, 钟益华, 王希哲, 等. 从单智能体到多智能体: 大模型智能体支持下的激励型学习活动设计与实证研究[J]. 华东师范大学学报(教育科学版), 2025, 43(5): 44–56.
- HUANG Changqin, ZHONG Yihua, WANG Xizhe, et al. J East China Norm Univ Educ Sci, 2025, 43(5): 44–56.
- [68] GHAFAROLLAHI A, BUEHLER M J. AtomAgents: Alloy design and discovery through physics-aware multi-modal multi-agent artificial intelligence[PP/OL]. arXiv (2024–07–13) [2026–04–08]. <https://arxiv.org/abs/2407.10022>.
- [69] 上海交通大学材料学院. DeepLight X AgentMat: 上海交大与小米联合发布轻合金大模型多智能体设计平台[EB/OL]. [2025–12–19]. <https://www.yzweekly.com/jsyy/10588.htm>.

作者贡献声明:

张林桐: 论文构思、设计研究方法和实验方案、撰写初稿

刘传宝: 调查研究、绘制图表

白 洋: 提供资源、获取资助、监督指导、审阅与修订

AI + Materials Science: Research Progress and Perspectives on Large Models in New Materials

ZHANG Lintong¹, LIU Chuanbao², BAI Yang²

(1. Basic Experimental Center for Natural Science, University of Science and Technology Beijing, Beijing 100083, China;

2. Institute for Advanced Materials and Technology, University of Science and Technology Beijing, Beijing 100083, China)

Extended Abstract

The field of materials science as a fundamental cornerstone for technological advancements and industrial transformation in society currently undergoes a profound paradigm shift propelled by artificial intelligence (AI). Conventional research and development (R&D) of materials, heavily reliant on empirical "trial-and-error" approaches and expert intuition, are inherently constrained by protracted R&D cycles and substantial financial costs. The advancement of the Materials Genome Initiative (MGI) has accumulated various large-scale scientific databases covering structure, property and process, but it remains a challenge that efficiently mines inherent law from these complex and multimodal massive data and integrates interdisciplinary information. The recent emergence and rapid evolution of large-scale foundation models offer a transformative solution to this critical bottleneck. Pre-trained on massive and diverse scientific corpora that include research literature, structural databases, and computational outputs, these models are evolving beyond mere pattern-recognition tools. They advance into sophisticated cognitive engines capable of

decoding specialized material languages, integrate physical laws across multiple scales, and ultimately restructure the entire material discovery and optimization workflow. This review systematically examines a pivotal role of Materials Foundation Models (MFMs) in this AI-driven revolution. It delves into their core methodological foundations, showcases landmark applications across key sub-disciplines, and outlines the trajectory leading toward a future of fully intelligent and autonomous materials discovery.

The significance of MFMs is fundamentally rooted in their capacity to establish a unified and comprehensive evaluative framework, which is structured around their core underlying training paradigms. This framework is primarily built on three synergistic pillars, including large-scale pre-training on structural and textual corpora, supervised fine-tuning for specific property regressions and reinforcement learning from physical feedback, which aligns model outputs with inviolable thermodynamic and kinetic constraints. This review deconstructs their capabilities and limitations *via* analyzing models through this tripartite lens. A detailed architectural analysis of pioneering vertical domain models, such as MatMind for inorganic non-metals and SteelBERT for metallurgy is given. We assess the specific mechanisms through which these models encode a deep domain-specific knowledge, including crystallographic symmetry rules, polymer chain topology dynamics, and solid-state reaction kinetics, into their latent representation spaces, thereby achieving a superior performance over general-purpose models. Furthermore, this review critically analyzes the sophisticated bridging mechanisms enabled by multi-agent frameworks. We examine how multi-agent systems like AtomAgents function as cognitive hubs, connecting and orchestrating high-throughput computational pipelines with experimental databases through unified natural language interfaces. These frameworks effectively manage the complex orchestration of diverse simulation tools and characterization equipment, thereby automating and accelerating the entire “design-simulate-characterize-optimize” cycle.

The research advances demonstrate a disruptive impact of these models across the materials R&D chain. MFMs significantly enhance design efficiency and prediction accuracy *via* suppressing “scientific hallucinations” through the integration of explicit physical rules and attention visualization techniques. In the realm of crystal discovery, the GNoME model effectively expands the library of stable inorganic crystals by ten times through an active learning closed-loop. For structure generation, models like CrystaLLM demonstrate an ability to capture atomic-scale symmetry breaking during phase stability predictions *via* deconstructing attention distributions over crystal descriptors. In specialized sectors, vertical models such as PolyLLM and ChipNeMo effectively bridge knowledge gaps *via* embedding domain-specific heuristics, while in the metallurgical sector, SteelBERT and PDGPT achieve a high accuracy in predicting steel properties *via* correlating process parameters with microstructural evolution. Moreover, multimodal frameworks such as MatterChat bridges a gap between natural language and atomic-scale features, where attention weights are automatically assigned to active-site orbitals, which is highly consistent with coordination field theory. The integration of agentic workflows further demonstrates that these foundation models can manage the entire cycle with an unprecedented precision.

Summary and Prospects MFMs decisively evolve from auxiliary computational tools into core engines for end-to-end intelligent R&D. They offer a unifying framework for the longstanding inverse design problem, enabling a systematic search of the MGI guided by target performance. Nevertheless, several challenges must be overcome to realize their full potential. Key limitations include the data scarcity for novel or extreme-condition materials, a need for even stricter integration of physical laws to ensure causal reasoning beyond correlation, and a “black-box” interpretability issue that can hinder trust in critical applications. Future development must focus on creating models with intrinsic causal reasoning capabilities, moving from statistical prediction to understanding the underlying mechanism of material features. This will be coupled with the integration of MFMs into fully autonomous laboratory platforms, where real-time perception from *in-situ* characterization and robotic control will enable closed-loop, self-driving experimentation. Ultimately, the convergence of AI and materials science steers a field from task automation toward high-level cognitive intelligence. This transition promises to fundamentally disrupt a conventional design logic, accelerating the discovery and deployment of advanced materials and ushering in a novel era of intelligent materials genome engineering.

Keywords materials science; artificial intelligence; materials foundation models; agent