

专题: AI 物质科学

大语言模型加速材料设计——从知识挖掘到智能设计的全链条赋能*

黄钰丹¹⁾ 夏琬钧²⁾ 杜俊梅¹⁾ 蒋渝¹⁾ 汪鑫¹⁾
陈元正¹⁾ 王红艳¹⁾ 赵纪军^{3)†} 郭春生^{1)2)‡}

1) (西南交通大学物理科学与技术学院, 成都 611756)

2) (西南交通大学图书馆, 成都 611756)

3) (华南师范大学物理学院, 广州 510631)

(2025年4月17日收到; 2025年6月3日收到修改稿)

随着人工智能技术的飞速发展, 大语言模型已经成为材料科学研究范式变革的核心驱动力. 本文系统地综述了大语言模型在材料科学全链条中的创新应用: 在知识发现与挖掘领域, 大语言模型凭借高效的信息检索和数据提取能力, 为材料研究提供了关键起点, 奠定了新范式的基础; 在材料设计与实验优化方面, 大语言模型通过跨尺度知识融合与智能推理, 能够揭示数据间的潜在关联, 在加速计算、合成设计、结构与性质预测、逆向设计等关键环节提供极具价值的解决方案, 大语言模型与自动化实验平台的深度融合, 实现实验流程的自然语言控制, 显著地提升了高通量实验的迭代效率. 研究表明, 大语言模型通过知识挖掘、知识推理与流程控制的三元协同, 正在重塑材料研发的全流程. 展望未来, 随着多模态感知与可解释性增强技术的发展, 大语言模型将推动材料科学研究进入新阶段.

关键词: 大语言模型, 材料设计, 智能任务, 合成与性质预测

PACS: 81.05.Zx, 81.20.-n

DOI: 10.7498/aps.74.20250497

CSTR: 32037.14.aps.74.20250497

1 引言

人工智能技术的迅猛发展, 为材料研发带来了前所未有的机遇, 正深刻变革着材料设计的理论与实践. 在人工智能的发展历程中, 机器学习 (machine learning, ML)、深度学习 (deep learning, DL) 以及大语言模型 (large language models, LLMs) 等技术不断演进, 为材料科学的研究提供了越来越强大的工具.

早期的 ML 方法主要依赖于手工特征提取. 例

如, 支持向量机通过人工设计的特征来划分数据类别, 随机森林则通过构建多个决策树来提高分类的准确性和稳定性. 这些方法在处理简单的线性关系和小规模数据集时表现出色, 但面对材料科学中复杂的非线性关系和高维数据时, 手工特征提取的局限性逐渐显现, 难以捕捉到材料性能的深层次规律. 随着 DL 的兴起, 自动特征提取机制成为可能. 深度神经网络, 如图神经网络, 能够自动从数据中学习到复杂的结构特征, 用于处理图结构数据; Transformer 架构则在处理序列数据方面表现出色, 能够捕捉到文本数据中的长距离依赖关系. 这些模

* 国家自然科学基金 (批准号: U2167217) 和四川学术成果分析与应用研究中心 (批准号: SCAA24-B19) 资助的课题.

† 通信作者. E-mail: zhaojj@scnu.edu.cn

‡ 通信作者. E-mail: csguo@swjtu.edu.cn

型通过自动学习特征,显著提高了对复杂体系的建模能力,为材料性能预测和结构分析提供了更强大的工具.近年来,LLMs的出现为材料科学带来了新的突破.以GPT-4(OpenAI),BERT(Google),DeepSeek等为代表的先进模型,通过深度神经网络架构对海量语料的学习,实现了对语言语法结构、语义逻辑的精准建模.LLMs不仅在自然语言处理(natural language processing, NLP)领域展现出强大的能力,还在跨模态理解与生成方面取得了显著的进展.它们能够将文本、图像等多种模态的数据进行融合,实现从知识挖掘到智能设计的系统性创新.

当前技术应用已经显现出重构材料研发全生命周期的可能性,催生出从知识挖掘到智能设计的系统性创新.LLMs在从知识挖掘到智能设计的全链条赋能主要体现在两方面,如图1所示:

1. 材料信息挖掘与知识发现.LLMs驱动的高效知识发现与挖掘是材料科学研究新范式的关键起点.

1) LLMs赋能的多模态信息检索技术,辅助研究者高效获取前沿且全面的知识,为材料研究奠定坚实基础.

2) 文本数据结构化是通过多种工具将信息之间建立起联系,命名实体识别(named entity recognition, NER)、提示词模板技术,研究者可以从海量复杂资料中精准且高效地提取关键数据信息,作为项目数据库内容,为后续研究任务提供丰富且准确的数据支持.

3) 构建知识图谱、知识库和创建智能问答系统是将结构化、半结构化和非结构化数据整合成一个系统性的图谱或存储信息的库,可帮助其快速获取全面的知识体系,从而加速研究进程.

2. 材料设计与实验优化.LLMs在材料科学研究新范式中的关键作用体现在跨尺度知识融合与智能推理两个维度.

1) 自动化实验室是LLMs赋能材料研究的延伸,它能帮助其快速获取全面的知识体系,从而加速研究进程.

2) 在材料设计的关键环节,如在加速计算、建模、合成设计、结构与性质预测、逆向设计等方面,LLMs凭借其强大的数据处理与推理能力,为材料创新提供了极具价值的创新解决方案.

3) 数据生成与实验迭代优化为研究人员的研究进程进一步加速.

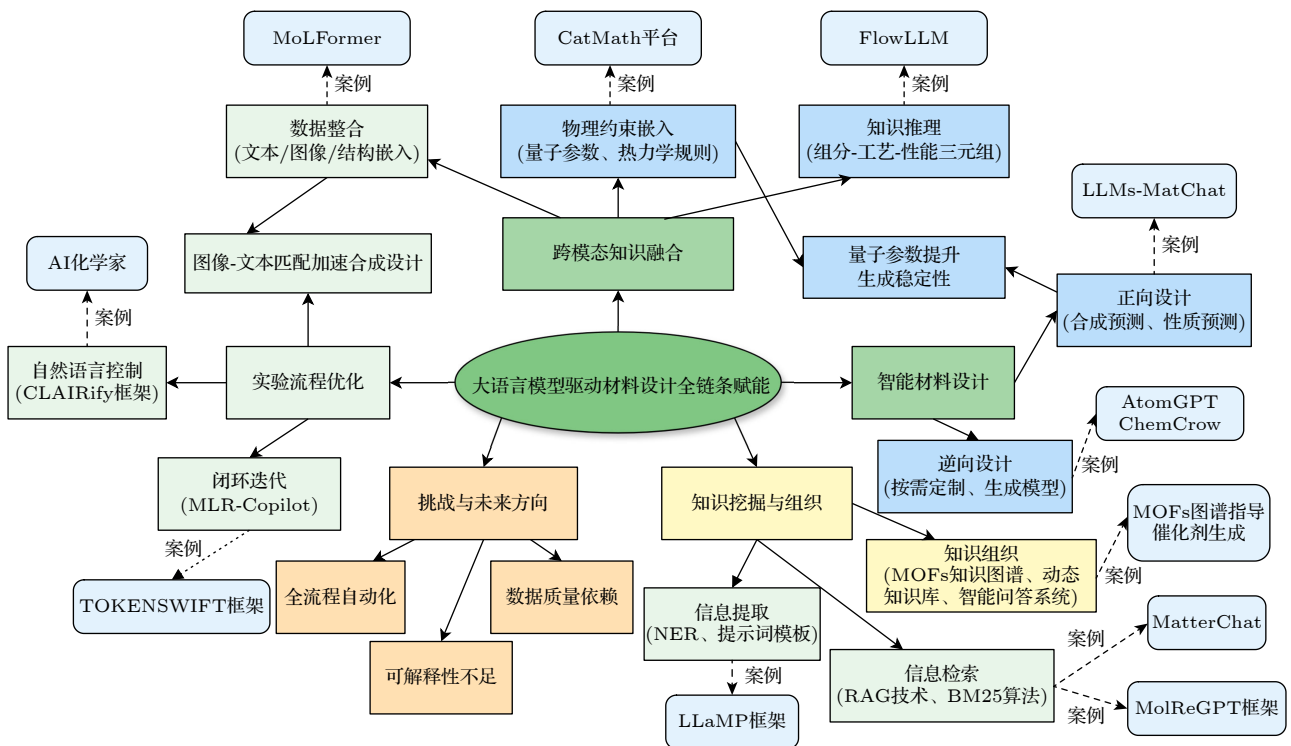


图1 大语言模型加速材料设计全链条赋能示意图

Fig. 1. Schematic diagram of the full-chain empowerment of material design accelerated by large language models.

本文旨在简要介绍 LLMs 从知识挖掘到智能设计对材料科学研究全链条的赋能, 加速材料设计研究的最新进展, 并对该领域的未来发展趋势进行展望. 研究发现, 其中知识挖掘环节与图书馆的学科情报服务高度契合, 各高校图书馆有望在其中发挥巨大作用. 总体而言, 虽然目前大多数 LLMs 在材料领域的应用仍处于探索阶段, 多数模型在处理相对简单、标准化的任务时表现出色, 但在面对复杂材料体系的深层次理解和精准设计时, 仍显得力不从心, 与此同时, 数据质量和数量的瓶颈也制约了模型性能的进一步提升. 但 LLMs 在材料科学领域已经展现出巨大的潜力, 能够从多个维度显著加速创新进程, 成为推动材料科学研究范式变革的核心驱动力. 展望未来, 通过数据融合、算法优化以及可解释性研究等技术提升, LLMs 有望突破当前的局限, 推动材料科学迈向新的发展阶段.

2 大语言模型加速材料信息挖掘与知识发现

在大数据时代, 材料科学领域的研究产生了海量的多模态数据, 包括文本、图像、音频等多种形式. 这些数据中蕴含着丰富的信息, 对于材料的设计与发现具有重要的指导意义. 数据驱动的知识构建旨在通过有效的技术手段, 从这些复杂多样的数据中提取有价值的知识, 为材料科学的发展提供有力支持.

2.1 材料信息检索与获取

2.1.1 LLMs 结合 RAG 在材料信息挖掘中的优化应用

检索增强生成 (retrieval-augmented generation, RAG) 是由 Facebook AI Research (FAIR) 团队于 2020 年首次提出的, 指出 RAG 能够利用检索机制从外部知识源 (如维基百科) 中获取相关信息, 优化 LLMs 的知识库, 提升其回答的精确性与丰富性, 弥补 LLMs 在知识更新和准确性方面的不足; 同时, LLMs 强大的语言生成能力可以将 RAG 检索到的知识更好地整合到生成的答案中, 使答案更加准确、具体和多样, 从而在知识密集型任务中取得更好的效果^[1]. 2024 年, Shi 等^[2] 通过结合 RAG 和 LLMs, 利用 BM25 算法从大规模数据中检索出与当前任务最相关的文本示例, 为 LLMs 提供精准的上下文信息, 从而显著提高信息检索的准确性和相关性. 在提取金属有机骨架 (metal-

organic frameworks, MOFs) 合成条件的任务中, 该方法在随机抽取的 123 个 MOFs 样本上, 将平均 F_1 性能从零样本学习的 0.81 提高到了 0.93, 提升了 14.8%. 同时, LLMs 强大的语言理解和生成能力充分利用检索到的高质量数据, 进一步提升信息提取的精度和效率, 实现了更优的检索性能和更可靠的结果. 在 MOFs 结构预测任务中, 与零样本 LLMs 相比, 该方法在随机森林等六种机器学习模型中均显著提高了预测能力, 平均决定系数提高了 29.4%. Luu 和 Buehler^[3] 通过微调和 RAG 技术得到了名为 Bioinspired LLM 模型, 能准确回忆生物材料信息并协助科研任务. 它基于大量相关文献微调具备推理和知识回溯能力, 在材料数据的结构相似性检索中, 该模型利用 LLMs 高效处理和理解复杂文献, 为研究人员提供准确且相关的检索结果, 加速生物启发材料的设计和发现.

2.1.2 结构相似性检索在材料数据中的应用

结构相似性检索是查找与给定对象结构相似对象的技术, 它能高效、精准地从大规模数据库中检索出与目标结构相似的材料, 显著加速新材料的发现与设计. LLMs 在这一过程中发挥着重要作用: 它能处理和整合材料数据, 理解用户检索需求, 优化检索过程, 分析并评估检索结果, 还能基于结果进行知识发现和创新, 辅助材料设计与优化, 最终生成相关报告与建议. 这贯穿了从数据预处理到检索及结果应用的整个流程, 显著提升了材料数据检索的效率、准确性和实用性, 为材料科学领域提供了强有力的支持. Li 等^[4] 提出了结合 LLMs 的上下文学习能力与分子相似性检索策略的 MolReGPT 框架. 该方法在分子转描述 (Mol2Cap) 和描述转分子 (Cap2Mol) 任务中, 利用检索的相似分子-文本对作为上下文示例, 指导 LLMs 生成高质量结果. 此外, Park 等^[5] 利用大型预训练化学基础模型 (如 MoLFormer) 生成分子结构的嵌入表示, 实现了材料数据中的结构相似性检索. 这些嵌入表示不仅能够用于小分子、聚合物和反应的结构相似性查询, 从而在复杂的研究任务中提供更丰富的上下文信息和更高效的材料设计支持.

2.1.3 多模态信息检索在材料数据中的应用

近年来, 多模态信息检索在众多领域广泛应用, 解决了多种复杂问题, LLMs 在这一过程中发挥着关键作用, 其强大的语言理解和生成能力, 不仅提升了文本语义检索的精准度, 还通过跨模态学

习与融合, 进一步增强了多模态信息检索的综合效能. 在现代材料研究中, 传统的单一数据类型检索方法已难以满足多源信息整合的需求, 多模态信息检索技术应运而生. 它打破数据类型壁垒, 将文本、图像、音频、视频等多种模态数据统一纳入检索范围, 更有效地挖掘数据潜在关联. Chiang 等^[6]提出的 LLaMP 框架, 通过整合文本、图像和结构数据等多种模态信息, 显著提升了材料属性预测的准确性. 与传统单模态检索和无多模态增强的预测方法相比, LLaMP 框架在多个性能指标上取得了显著改进. 例如, 在预测材料的体积模量时, LLaMP 的平均绝对误差从传统方法的约 40 GPa 降低到 14.57 GPa; 在预测形成能方面, LLaMP 的平均绝对误差也显著低于传统方法. LLaMP 框架采用 RAG 范式, 通过检索机制从外部知识源获取相关信息, 为 LLMs 提供更精准、更丰富的知识支持. 此外, LLaMP 引入了多层次的推理与行动代理, 实现动态且递归的交互. 这种多模态数据的整合与交互机制, 为材料科学中的复杂问题提供了新的解决方案, 推动了材料设计的智能化发展. Tang 等^[7]将 LLMs 与材料结构数据相结合, 开发出一种多模态模型 MatterChat. 该模型通过桥接模块整合材料的原子结构和文本信息, 利用 LLMs 处理和理解材料相关的文本描述, 同时通过结构相似性检索技术捕捉材料的结构特征, 从而实现对材料性质的准确预测和复杂的人机交互. 针对材料科学中多模态数据分散异构、关键信息提取困难及跨模态语义割裂等问题, 李长泰等^[8]提出基于 LLMs 的多模态信息检索技术, 通过整合文本、图像等多模态数据, 实现全面精准的材料信息检索与分析.

2.2 材料知识提取与结构化

在数据驱动的知识构建中, 从海量文献和数据里提取有用知识并转化为结构化形式是关键步骤, 这不仅提升知识的可访问性和可利用性, 还为后续材料设计和性能预测筑牢数据根基. 通过将数据结构化、NER 和提示词模板优化运用于 LLMs, 可高效实现知识提取与结构化, 有力支持数据驱动的知识构建.

2.2.1 文本数据结构化

文本数据结构化是将非结构化或半结构化文本转化为结构化数据, 使信息与信息之间的关系更为明确, 这对材料信息库建设意义重大. 近年来,

诸多研究利用 LLMs 提升此过程效率, 如 Gupta 等^[9]开发的 MatSciBERT 模型, 这是一个专门针对材料科学领域的语言模型, 旨在加速文献中的信息提取和知识挖掘. 该模型基于 BERT 架构, 使用涵盖多个材料科学领域约 2.85 亿单词的材料科学语料库进行预训练. 预训练采用掩码语言模型任务, 并结合动态整词掩码、去除下一句预测损失等优化策略. 在微调阶段, MatSciBERT 在 NER、关系分类和论文摘要分类等下游任务中表现出色, 通过条件随机场层和实体标记器等技术提升性能. 与 SciBERT 和 BERT 等通用模型相比, MatSciBERT 在多个性能指标上显著提升, 例如在 SOFC-Slot 数据集的 NER 任务中, Macro-F1 分数比 SciBERT 提升约 6.3%, 比 BERT 提升约 8.9%. 为解决传统方法受限于高维参数空间探索效率低和文献知识利用不足的问题, Lai 等^[10]提出了一种基于 LLMs 的文本数据结构化方法, 该方法利用 LLMs 从科学文献中自动提取催化剂合成参数, 构建多维度化学空间, 并结合贝叶斯优化进行多目标参数寻优. 2024 年, Thway 等^[11]通过设计特定的提示集, 使 LLMs 能够从大量文本中提取关键的固态合成信息, 并将其结构化为表格格式. Yu 等^[12]通过 LLMs 从非结构化的科学文献中提取结构化数据, 包括获取非结构化文本、数据清洗、初步筛选、初步数据提取、验证和应用于下游任务. Durmaz 等^[13]通过结合本体知识库与 LLMs 实现了文本数据的结构化, 他们利用 LLMs 对文本进行初步理解和分析, 为标注者提供参考和提示, 从而提高标注效率和质量. 同时, LLMs 能够解析复杂句子结构, 识别实体关系的潜在模式, 为关系提取和知识图谱构建提供有力支持. 为了实现材料科学文献中显微图数据的文本数据结构化, Lei 等^[14]利用 LLMs 对文献中的标题、摘要和图注等文本信息进行筛选和标注. 通过结合视觉语言模型对显微图进行图像分析, 研究者成功生成了一个包含 842 张显微图的结构化数据集, 每张显微图均附有仪器和材料标签. 这些方法大大提高了信息提取的效率和准确性, 为材料信息库的建设提供了有力支持.

2.2.2 命名实体识别在材料信息提取中的应用

NER 是 NLP 中用于提取材料关键信息的基本任务, LLMs 通过预训练学习了丰富的语言知识和语义信息, 为 NER 任务提供了强大的支持. LLMs

能够准确识别材料名称、成分等实体,并在不同来源和风格的材料科学文献上取得良好效果.同时,NER技术能够增强LLMs在材料信息提取中的表现,提升信息抽取精度,实现联合抽取,构建更完整的知识结构,并促进领域知识融合. Weston等^[15]开发了一种基于NER和实体归一化的技术,用于从材料科学文献中提取信息.该技术通过预处理、分类器筛选和神经网络模型,从327万篇摘要中提取了8000万个实体,并将其归一化为标准形式.与传统方法相比,该技术具有更高的准确率和召回率,显著提升了材料科学文献的检索效率和知识利用效率.但随着LLMs的发展,这一领域的研究取得了更大的突破.如Zeng等^[16]开发的ChatMol模型,它在NER应用于材料信息提取中发挥了重要作用.ChatMol利用LLMs强大的语言理解和表示能力,能够准确识别材料科学文献中的材料名称、化学物质等命名实体,并减少误识别^[16].李长泰等^[8]指出LLMs在材料科学领域的NER中发挥着关键作用.通过预训练与微调LLMs能够精准识别并抽取材料科学文献中的关键实体信息,极大提升了信息提取的效率与精准度,为材料知识图谱构建及新材料研发提供了强有力的支持.针对热电材料领域的大数据,Jia等^[17]提出通过LLMs的NER技术,从非结构化文献中提取热电系数、掺杂元素和晶格参数等关键指标,并结合多模态数据融合方法构建标准化热电材料数据库,显著提升了AI模型训练数据的质量和覆盖范围.Polak等^[18]在研究中指出LLMs在NER应用于材料信息提取过程中发挥了关键作用,通过对话式的交互和信息保留能力,LLMs能够精准识别并提取材料名称、数值和单位等关键信息.同时,模型通过后续的验证问题引入不确定性,减少错误和幻觉数据的产生,从而实现从科研文献到结构化材料数据库的高效转化.另外,Foppiano等^[19]评估了LLMs在材料科学文献中的NER任务上的表现,特别是对材料和物理性质的识别.研究表明,MatBERT在材料科学领域表现更优,GPT模型经过微调后在部分数据集中NER任务精度超传统模型,但在LLMs零样本提示下表现得有限,而经过微调后在关系抽取任务表现得更优异.为了解决材料化学领域复杂知识提取难题,Dagdelen等^[20]提出了一种基于微调LLMs的联合NER与关系抽取方法,该方法通过将科学文本映射至用户定义的结构化模式实现了实体自动纠错,且在信息提取等任务中显

著优于传统模型.以上研究成果为分子生成与材料科学相关的信息处理提供了切实可行的技术路径.

2.2.3 提示词模板在材料信息抽取中的优化应用

提示词模板是一种用于引导LLMs生成特定内容的文本框架,由固定部分和可变部分组成,固定部分包含通用引导语句或关键词,可变部分是用占位符等来表示的供用户依情况填充的部分.Zheng等^[21]通过精心设计的提示词模板,利用ChatGPT实现了从科学文献中高效提取MOF合成条件的目标.他们开发的ChatGPT化学助手(ChatGPT chemistry assistant)通过优化的提示词模板,引导ChatGPT准确识别并提取文献中的关键实体,如金属源、有机连接体、溶剂及其用量等,并将这些信息结构化为表格格式^[21].这一方法不仅显著地提高了信息提取的精度和效率,还为后续的材料设计和性能预测提供了高质量的数据支持.为了提高材料信息抽取的准确性和效率,Yang等^[22]设计了六种提示词模板,通过优化提示词模板的应用,充分发挥LLMs的强大语言理解和生成能力,使其能够更精准地理解和处理材料科学文献中的自然语言文本,从而实现材料信息的智能抽取和整理.2024年,时宗彬等^[23]凭借提示词模板来实现材料信息抽取,该提示词模板由前缀、指令、示例和输入文本组成,旨在减少LLMs输出幻觉内容,同时提高结果的精确度和可阐释性.模板通过明确的指令和示例引导LLMs生成所需的信息,设计准则包括尽量减少LLMs输出幻觉内容,同时确保结果易于解释.

2.3 材料知识组织与问答

2.3.1 知识图谱在材料信息组织中的应用

LLMs在材料科学知识图谱构建中发挥着关键作用,它能够高效处理海量材料科学文献、实验报告及数据库等多源异构数据,依托其卓越的NLP与语义理解能力,精准识别并提取材料的成分、结构、性能、制备工艺等关键知识要素.同时,LLMs可依据预设规则和逻辑,将这些知识要素映射为知识图谱中的节点与边,自动化构建知识关联,极大提升知识组织效率与质量,为材料科学知识图谱的完善与应用提供强大助力.为了解决材料科学中的复杂问题并推动材料信息的高效组织与应用,Buehler^[24]运用LLMs通过强大的NLP能力,从

大量材料科学文献中提取关键信息,包括材料的性能、制备方法、应用领域等,并识别不同概念之间的关系,如材料与性能、材料与应用之间的关联,这些信息和关系被整合进知识图谱构建出一个结构化的材料知识网络.同时,LLMs利用知识推理能力,基于已有知识图谱预测新材料的性能和应用,通过交互式知识获取,用户可以用自然语言进行提问,模型从知识图谱中检索并返回答案,实现知识的高效利用和扩展.Zia等^[25]利用LLMs从材料科学文献中自动提取信息,构建知识图谱以组织材料无损检测知识.LLMs精准识别并提取文献中的关键信息,包括材料类型、退化机制、物理变化及对应的无损检测方法等,并将这些信息分类整理,转化为知识图谱的节点和关系.Corlatescu等^[26]详述基于逐句文本解析的AMoC v4.0知识图谱构建流程.步骤包括:从文本中提取概念和属性节点,利用ChatGPT生成边(连接两个节点的关系或连接)以连接这些节点,然后将生成的节点和边添加到知识图谱中.对于后续句子,模型会保留前一句的活跃概念、属性和边,再构建新的知识图谱.针对框架材料(MOFs, COFs, HOFs)领域文献分散、知识整合低效的问题,Bai等^[27]通过LLMs自动化处理超10万篇文献摘要,并提取实体间关系,构建了包含253万个节点和401万关系的知识图谱.经测试表明,该图谱能精准挖掘隐含信息,并赋能LLMs开发专业问答系统,显著解决传统模型“幻觉”问题,为材料设计、趋势预测及跨学科研究提供了结构化知识库与智能化工具.

2.3.2 知识库在材料信息组织中的应用

知识库作为组织知识资产的核心载体,通过整合分散信息资源、打破部门壁垒,以结构化向量数据库驱动高效查询与智能交互,实现决策加速、流程优化及创新赋能.其构建过程包括目标设定、数据采集、文本处理等环节,最终通过精准文档匹配与提示工程实现知识服务闭环,该过程强调数据质量管控、模型安全加固与持续迭代优化,确保组织知识资产的长效积累、动态更新与价值转化,成为支撑数字化转型与智慧运营的关键知识基础设施.贺强等^[28]借助LLMs强大的NLP能力,高效地从大量维修案例文本中提取关键信息得到了复合材料结构适航维修知识库.LLMs通过语义理解与分析,精准地设计语义检索索引和属性检索索引并制定相似性评价方法,实现案例的快速准确检索与匹

配.同时,LLMs定制数字化维修模板,将检索结果与模板关联,为维修方案设计提供科学依据,提升维修效率与质量.陈慧琳^[29]凭借LLMs强大的NLP能力加速材料信息知识库构建,首先从海量文献和网页文本中精准提取材料的关键信息,如性能参数等,并将其转化为结构化的数据格式以便于知识库的高效存储和管理;同时,LLMs能够深度理解材料领域文本的语义,准确识别专业术语和概念之间的关联,从而将分散的材料信息整合成系统化的知识体系,自动建立不同材料知识之间的复杂关联,如材料的相似性等关系,极大地丰富了知识库的内容和语义层次;此外,LLMs还可以作为智能辅助工具在知识库的更新和维护过程中,持续从新的材料研究成果和数据中学习,及时发现并补充知识库中的缺失信息,确保知识库的时效性和准确性.Ye等^[30]在材料科学运用LLMs,通过NER、关系抽取和实体解析从非结构化文献中精准提取材料信息并转化为结构化三元组,构建了包含162605个节点和731,772条边的材料知识库.该知识库整合分散知识,支持复杂关系分析和潜在联系预测,为材料科学研究提供可靠数据支持和创新灵感,加速新材料发现与开发.针对全固态电池材料研发的迫切需求,Yang等^[31]通过集成LLMs的文本挖掘与结构化能力,动态更新固态电解质的合成参数、离子电导率及稳定性数据.LLMs在此过程中不仅从海量文献中自动提取关键实体(如电解质成分、制备工艺),还通过知识推理生成材料性能的潜在关联规则,为知识库的持续优化提供智能化支持.由松山湖材料实验室开发的材料科学领域AI工具MatChat AI Agent是一款基于LLMs和增强搜索技术,提取了28万余篇材料科学、物理、化学论文中的知识,形成了结构化的知识库.该工具通过对话形式为用户提供了一种更精准的材料科学知识支持,同时避免了通用LLMs中的幻觉问题,确保信息的可靠性和学术性.MatChat AI的发布标志着材料科学研究向智能化、数字化方向迈出了重要一步^[32].以上知识库的构建使得研究人员能够更容易地发现和利用材料之间的复杂关系,从而在材料设计、合成预测、结构与性质预测等方面提供创新解决方案.

2.3.3 智能问答在材料信息组织中的应用

智能问答系统作为人机交互的关键桥梁,在材料科学领域展现出巨大潜力.它基于LLMs构建,

具备强大的 NLP 和生成能力,能够通过预训练学习海量文本数据中的语言规律和模式,并借助指令微调与强化学习等技术手段,针对特定领域的专业需求进行优化,为材料科学研究提供精准、高效的问答服务.如 Buehler 等^[33]通过微调和优化 LLMs 使其能够准确理解材料科学领域的专业问题,并从海量知识库中检索相关信息. LLMs 不仅负责解析用户提问将其转化为对知识库的有效查询,还利用其强大的语言生成能力将检索到的复杂信息整理成逻辑清晰、专业准确的回答.此外, LLMs 通过不断学习和吸收材料科学领域的最新研究成果,确保知识库的更新与扩展,从而提升整个问答系统的性能和实用性,使其能够更好地支持材料设计和分析任务. Cai 等^[34]研究了 ChatGPT 4.0 在水泥基复合材料领域的智能问答的应用,通过设计一系列难度递增的问题,评估其在材料信息组织方面的表现. LLMs 在智能问答系统构建中发挥了重要作用,它依托强大的 NLP 能力实现对材料信息的高效组织与精准检索,为研究人员和工程师提供初步的材料配方指导、裂纹分析及碳排放计算等服务,推动了水泥基复合材料领域的智能化发展.类似地,任海玉团队^[35]提出的智能问答系统是基于 LLMs 的核心能力,通过整合 RAG、知识图谱和智能代理技术搭建而成的.该方法有效地解决了传统问答系统在复杂语义理解、多层推理、知识实时更新等方面的局限性,显著减少了 LLMs 的“幻觉”问题. Bran 等^[36]提出的智能问答系统在材料信息领域展现了创新性应用,该系统借助 LLMs 的强大 NLP 能力能够精准理解用户提出的复杂问题,并通过与材料数据库和知识图谱的深度整合检索出与问题高度相关的信息. LLMs 在系统中不仅充当信息的检索工具,更作为智能分析的核心引擎,对材料的特性、应用范围以及合成方法等进行逻辑推理和知识整合.它能够处理多轮对话,根据用户后续的提问或澄清不断调整和优化答案,确保信息的准确性和针对性. Ansari 等^[37]将智能问答系统用于新材料发现,借助 GPT-4 和 Claude 3.5 Sonnet 从文献获取知识,通过人机协作,让科学家反馈、约束,引导 AI 更有效探索分子设计.为了解决材料设计中数据分散和知识碎片化的问题,并实现高效的知识整合与应用, Zheng 等^[38]通过 LLMs 从海量科学文献中提取材料合成参数和性质信息,构建了机器可读的数据库,为材料设计提供了丰富的数据支持.基于

图理论开发了知识图谱,整合材料的组成、结构和性能信息,实现了智能问答系统.该系统能够根据用户需求快速推荐合成路径和条件,显著提升了材料设计和发现的效率.

3 智能材料设计与实验优化的运用

基于第 2 节构建的多模态材料知识库、知识图谱与智能问答系统,材料研发正从经验驱动转向数据-知识双轮驱动的智能范式.知识图谱通过实体关系网络(如“材料组分-工艺参数-性能指标”三元组)为 LLMs 提供了结构化推理框架:在合成路径设计中, LLMs 可通过图谱中的反应热力学关联推理出可行反应路径;在实验流程优化中,跨模态检索技术可快速匹配历史实验数据与新需求间的结构相似性,指导工艺参数迭代.这种知识嵌入机制使 LLMs 突破纯文本处理的局限,将材料基因数据、工艺约束与性能目标深度融合,为后续的自动化实验等场景提供兼具物理意义与工程可行性的解决方案.

3.1 自然语言驱动的实验自动化

自然语言处理技术在实验机器人中的应用显著降低了操作复杂度,使用户能够轻松使用这些系统.例如, Yoshikawa 等^[39]在 2023 年提出的 CLAIRIFY 框架,他们通过将自然语言指令转换为机器人操作手可执行的计划,实现了化学实验的自动化,如图 2(a).在实际的材料合成实验场景中,这种融合展现出巨大优势.比如在制备新型陶瓷材料时,研究人员可以用自然语言向配备了 NLP 系统的实验机器人下达指令,机器人迅速解析指令,精准操控机械臂完成一系列精细操作.在球磨过程中,若遇到粉末结块等异常情况,机器人还能通过语音及时向研究人员预警,等待进一步指示,确保实验顺利进行.这种 NLP 技术赋能的实验机器人,有效地提升了实验操作的准确性与效率,降低了人为操作误差,为材料研发开辟了全新的智能化路径,极大地推动了材料科学的创新发展.

自动化实验系统的核心在于能够自主执行复杂的实验流程.近年来, LLMs 与材料化学的相关研究展示了全自动化学实验室的潜力.2022 年, Zhu 等^[40]介绍了一个 AI 化学家,如图 2(b),它能够自动读取文献、提出假设、设计实验计划、执行自动

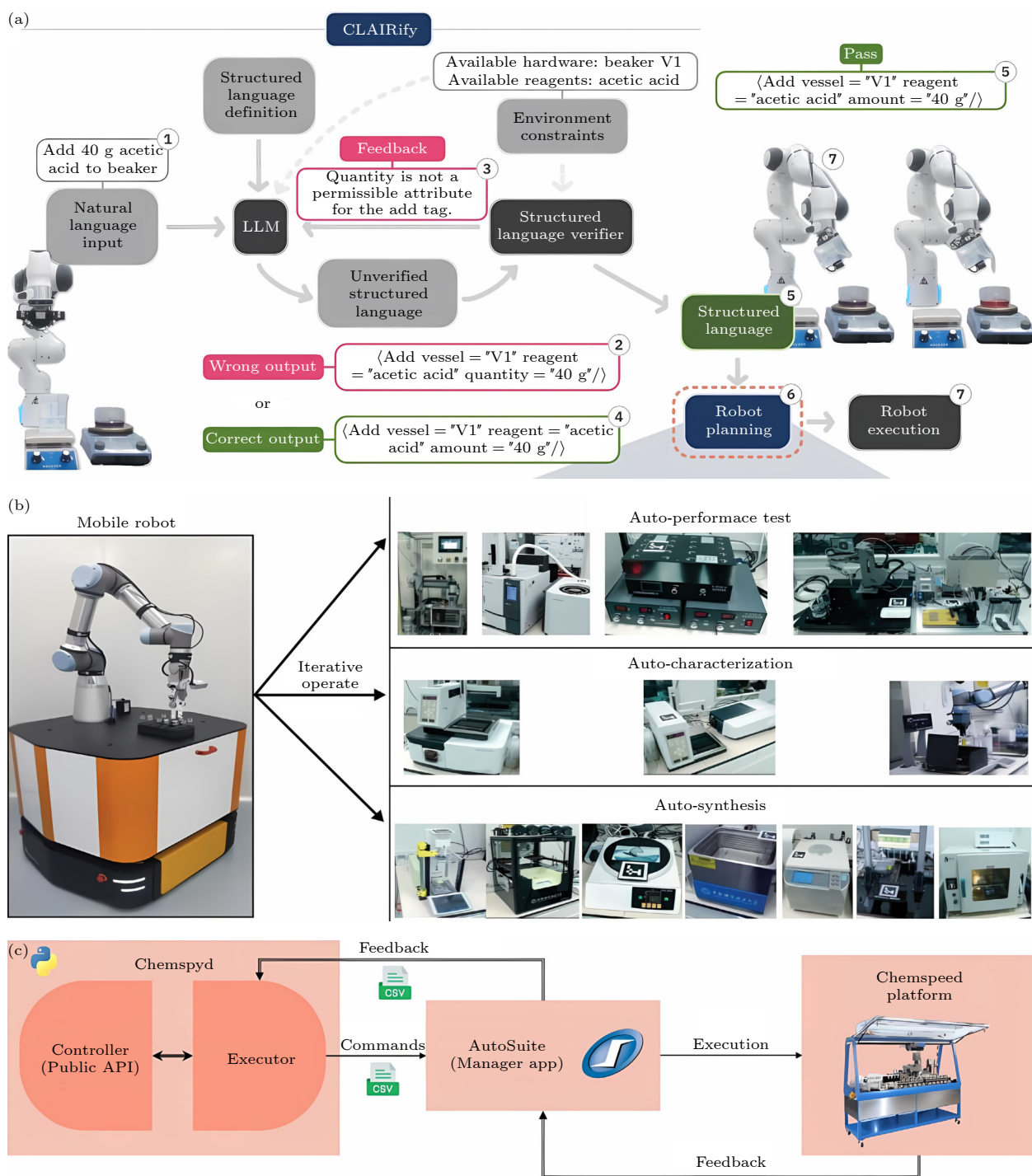


图 2 (a) CLAIRify 框架: 基于 LLMs 的 NLP 模块。LLMs 采用输入①; 结构化语言定义和资源约束, 生成未经验证的结构化语言②; 输出结果由验证者检查并通过反馈传递给 LLMs③; LLMs 产生的输出通过验证器④; 将正确的输出⑤; 传递给任务和运动规划模块⑥; 生成机器人轨迹⑦^[39]; (b) 自主移动机器人研究各个系统的工作流程和每个模块的功能^[40]; (c) Chemspeed 通过 Executor 与 AutoSuite 进行通信, Executor 读写共享的 CSV 文件, 提供了一种标准的通信方式, 这种通信方式是人类可读的, 并且由 Python 和 AutoSuite 共同支持^[42]

Fig. 2. (a) CLAIRify Framework: NLP Module Based on LLMs. The LLMs take input ①; structured language definitions and resource constraints are used to generate unverified structured language ②; the output is checked by a verifier and fed back to the LLMs via feedback ③; the output generated by the LLMs passes through the verifier ④; the correct output ⑤ is passed to the task and motion planning module ⑥; generating robot trajectories ⑦^[39]. (b) Workflow of various systems in autonomous mobile robot research and the functions of each module^[40]; (c) Chemspeed communicates with AutoSuite via the Executor, which reads and writes shared CSV files, providing a standardized communication method that is human-readable and supported by both Python and AutoSuite^[42].

化操作、分析实验数据、训练 ML 模型并提出新的假设。随后 2024 年,为了解决传统自动化实验室在化学合成中面临的设备定制化、单一表征技术依赖以及决策算法局限性等问题,实现从合成到分析再到决策的全流程自动化, Dai 等^[41]开发了一个模块化的自主实验平台。该平台利用移动机器人将现有的实验室设备整合在一起,无需对设备进行大规模改造或垄断使用。通过结合液相色谱-质谱联用仪和核磁共振波谱仪等多种分析技术,该平台能够处理复杂的多步骤合成反应并通过启发式决策算法对反应结果进行评估,选择成功的反应进行后续的放大和功能测试。Seifrid 等^[42]引入了一个 Chemspeed Technologies 的 Chemspyd 开源的 Python 接口,用于操作 Chemspeed 自动化平台,如图 2(c)。该系统支持与 LLMs 的集成,提供了自然语言界面,进一步简化了实验操作流程。

此外, LLMs 如 GPT-3 等在实验中的应用展示了其在信息提取、任务规划和人机交互方面的潜力,尤其是在自动化和智能化方面。例如, Zheng 等^[43]提出了一种将 ML 与 LLMs 结合的方法,旨在加速电化学 C—H 氧化反应的探索和优化过程。利用 ML 开发了反应性预测模型和选择性预测模型,使用 LLMs 进行语义分析,从海量文献中自动筛选出相关的反应数据,除此之外, LLMs 还被用来自动生成 ML 代码,简化了反应条件优化的过程,使得化学家能够更高效地进行实验设计和数据分析。Ruan 等^[44]开发了一种基于 LLMs 的端到端化学合成开发平台 (LLM-RDF),通过整合文献检索、实验设计、硬件控制、数据分析等六大智能代理模块,实现了从文献挖掘到产物纯化的全流程自动化。其自然语言交互界面降低了自动化实验门槛,将传统依赖专家经验的试错过程转化为智能闭环系统,为化学合成开发提供了高效、低成本的解决方案,推动了人工智能与实验科学的深度融合。

总之,实验机器人和机械臂智慧实验平台的发展为化学和材料科学的研究带来了新的机遇。通过结合 NLP、LLMs、自动化学实验系统和 ML 算法,研究人员能够更高效地进行实验设计和数据分析,推动科学发现的进程。例如 Hatakeyama-Sato 等^[45]利用液体处理装置和通用机械臂合成聚酰胺酸颗粒,通过集成摄像头和多模态 LLMs 持续监控和记录实验过程,自动生成详细的实验笔记。为了解决自动化化学实验室中复杂化学实验的高效执行问

题,实现多机器人协同工作以提升实验效率, Zhou 等^[46]提出了一种多机器人多任务调度系统。该系统通过约束编程方法优化任务分配,协调多台机器人和实验站的运作,实现了复杂化学实验的并行执行。在实际应用中,该系统针对四种不同化学实验进行调度,这些实验包括制备石墨相氮化碳 (g-C₃N₄)、合成五元金属-有机高熵催化剂 (MO-HECs)、研究二元层状双氢氧化物 (LDHs) 的尿素氧化性能以及探索小檗碱氯化物 (BBR) 的聚集诱导发射 (AIE) 特性。结果显示,与顺序执行相比,总执行时间减少了近 40%。未来的研究应继续关注这些技术的集成和优化,进一步提高实验自动化系统的灵活性和可靠性。

3.2 大语言模型在材料设计中的应用

3.2.1 合成预测与路线设计简化材料发现流程

人工智能可借助文本挖掘技术和先进的机器学习算法从海量的科学文献中提取材料合成所需的数据,进而精准预测合成特定材料的相关数据。传统方法依赖于研究人员手动收集和整理实验数据,所以需要大量的人力物力,而通过 ML 模型, LLMs 可以预测不同参数组合的实验结果,优化实验设计、提升科研效率。无机材料合成预测以晶体结构生成为核心,依赖材料基因组数据库中的晶格参数、空间群和相稳定性数据。以 CrystaLLM 为代表的模型采用自回归语言建模技术,能够生成有效的晶体结构,并通过蒙特卡罗树搜索算法进一步提高生成质量^[47]。为证明人工智能驱动平台对加速材料发现的可行性, Szymanski 等^[48]利用 LLMs 结合材料基因组计划和谷歌 DeepMind 的相稳定性计算数据,构建 ML 模型预测新型无机材料的合成路径。为提高生成稳定材料的效率,并减少后续处理计算的成本, Sriram 等^[49]于 2024 年在文章中专注于生成无机晶体材料的 FlowLLM 是通过结合 LLMs 和 Riemannian Flow Matching (RFM),能够生成新颖且稳定的晶体材料。Jia 等^[50]详细描述了 LLMatDesign 框架,该框架利用 LLMs 进行无机材料设计,包括合成预测、结构和性质的预测。通过迭代过程, LLMatDesign 能够在有限的训练数据下实现新材料的设计。这表明 LLMs 能够帮助科学家构思新的合成策略,并通过理论计算或模拟来验证这些假设的可行性。Liu 等^[51]通过整合 LLMs

的多模态数据解析能力与 ML 算法, 实现了热催化和电催化材料的高效性能预测与合成路径优化, CatMath 平台利用 LLMs 从海量文献中提取催化剂活性位点与反应条件的关联规则, 并结合知识图谱构建催化反应网络的物理约束, 显著提升了逆向设计中的候选材料筛选效率. 此类技术印证了 LLMs 在跨尺度知识融合与数据驱动的合成路径生成中的核心作用, 为复杂催化体系的创新提供了新范式. Zhang 等^[52]的研究指出, LLMs 可以使用多模态方法整合和分析对应的数据源, 辅助识别高效的电催化剂材料, 从而为新材料的合成提供指导. 2025 年, Slautin 等^[53]通过整合 ML、自动化实验平台 (如 A-Lab) 与扫描探针显微镜多模态表征技术, 预测了材料合成路径、功能性能与结构-性能关系. 实验验证显示, A-Lab 在 17 天内成功合成 41 种新材料, 组合库探索效率提升 50%, 量子点合成速率提高 10 倍, 为高通量材料设计提供了闭环优化范式. 传统材料合成路径设计长期受困于工艺参数的高维组合爆炸问题, LLMs 通过数据驱动的方式可以将参数搜索空间压缩, 快速给出智能分析和建议. 例如 Su 等^[54]用 ML 和 LLMs 通过实验或理论反馈指导实验设计. 通过结合 ML 算法, 可以更有效地探索可能的反应路径, 找到最佳的合成方案, 如图 3(a). 这种做法不仅避免了传统的筛选催化剂的试错法费时又费力问题, 又减少研发成本, 并且加快对催化剂寻找进程. 又例如, 中国科学院研发的专注于无机材料合成路径预测任务的 LLMs-MatChat 取得了显著成果. 他们从 400 多万篇论文中提取 35675 个无机材料固相反应合成过程, 精心处理了 13878 条高可信度的合成路径描述数据, 对开源大语言模型 LLaMA2-7B 进行微调训练. 经实验验证, 该模型在预测合成复杂的无机材料时, 性能表现超越 ChatGPT, 展现出大模型在国内材料领域的创新潜力与应用空间^[55]. 在催化领域, Digital Catalysis Platform (DigCat) 通过整合 LLMs 与大数据技术, 实现了催化剂设计与合成路径的智能化预测^[56]. 例如, 其利用 LLMs 从海量文献中提取反应条件与性能的关联规则, 并结合 ML 模型优化合成参数, 显著缩短了新型催化剂的研发周期. 在纳米酶催化活性的研究中, 传统方法可能需要大量的实验来探索不同材料和表面修饰对催化性能的影响, 耗时且成本高昂. 在 2024 年, Razlivina 等^[57]通过由 AI 驱动的包含 1210 个实验样本的纳米酶数据库, 根据特征选择描述符进行训练, 如

图 3(b). 这种基于集成学习算法的 ML, 通过设计实验参数预测纳米酶催化活性. Chen 等^[58]还建立了 LLMs 增强的 CO₂ 还原电催化剂及其合成程序的语料库, 如图 3(c), 有助于系统地理解和优化实验条件.

3.2.2 结构与性质预测助力材料发现

结构预测是材料科学中的一个关键领域, 旨在通过计算方法预测新材料的晶体结构和相变行为. 从而能够根据目标材料的结构与性质, 逆向重构出其可行的合成路线与必要条件. 通过将 LLMs 生成的文本与现有的科学文献相结合, 研究人员可以快速识别潜在的结构候选者, 进而推导出具备实验可操作性的研究假设.

2024 年 Liu 等^[59]提出了一种结合 LLMs、提示工程和 DL 的工作流程, 用于解决材料分类问题, 利用合金成分的文本描述和预训练模型提高预测准确性, 工作流程如图 4(a) 所示. 在结构预测中, 人们很关注 LLMs 是否能有效地捕捉微观结构中的复杂关系和模式, 在处理大规模粒子系统时, 能否仍然保持高效和准确. Chen 等^[60]提出了一种名为 MeLM 的多模态力学语言模型, 该模型处理复杂结构预测任务时表现出色, 能通过多种机制和编码方式预测微观结构, 工作流程如图 4(b). Wang 等^[61]利用 ChatGPT 辅助结构预测和材料设计, 结合文献提出新分子并进行验证, 工作流程如图 4(c) 所示.

性质预测需要结合结构信息和物理化学原理, 建立材料晶体结构与电子态同其物化性质之间的构效关系, 进而实现性能预测. 基于量子力学的第一性原理计算为材料性质预测提供了严格的理论框架. 虽然该方法在预测精度上具有无可争议的优势, 但其计算复杂度随体系规模呈指数级增长, 这一固有特性严重限制了其在扩展体系中的应用. LLMs 在性质预测中发挥着越来越重要的作用, 精确的预测模型可基于海量实验数据集构建而成, 利用数据驱动的方法重构材料的构效关系.

性质预测在多个领域都有应用, 涵盖理化性质、量子化学性质、生物活性及毒性等关键领域. 通过消息传递神经网络、图注意力网络、深度卷积模型等算法, 结合分子描述符、SMILES 编码、三维图结构等多模态特征化方法, 实现了实验成本降低与预测精度提升^[62]. 在合金材料领域, Liu 等^[59]在 2024 年提出了一种结合 LLMs、提示工程

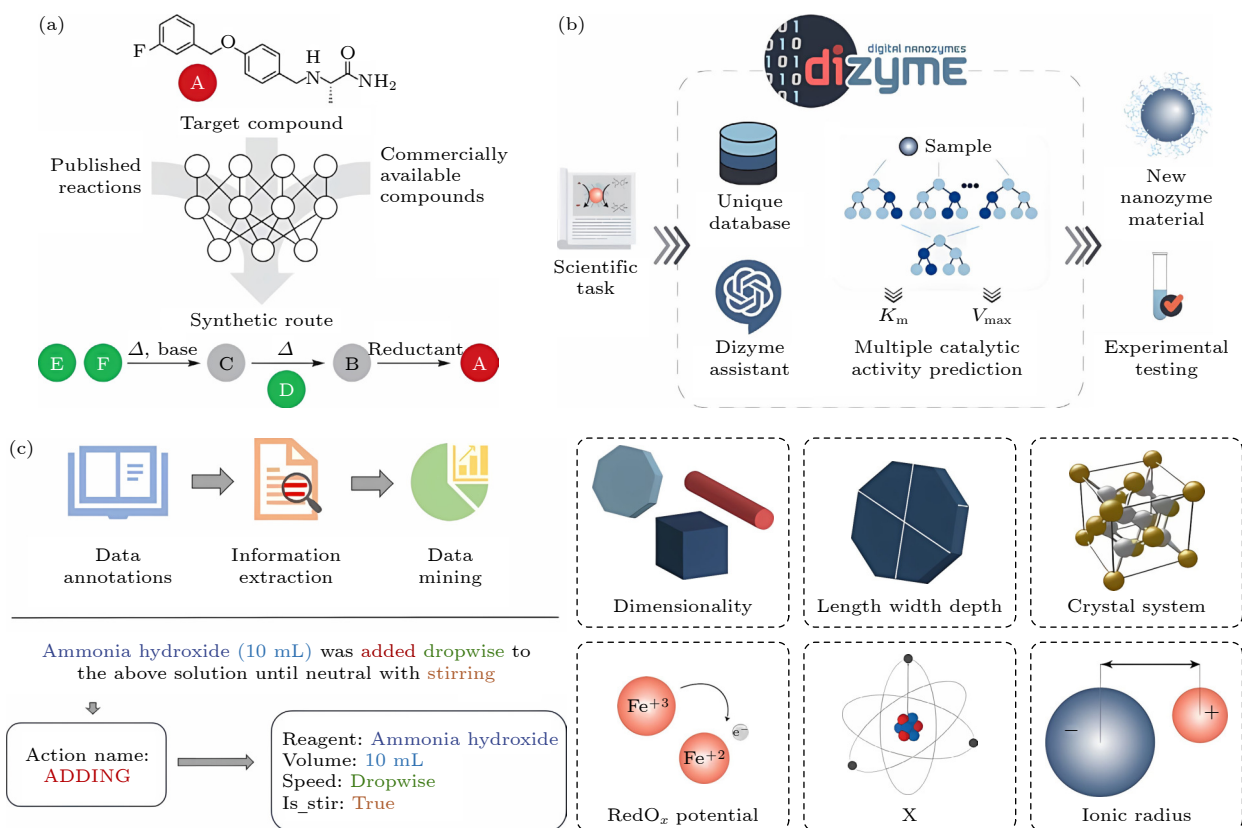


图 3 (a) 流动化学系统集成的多种反应模块和在线分析工具形成了连续的合成路径^[54]; (b) 上部分为 DiZyme 工作流程: 从制定科学任务到发现新的纳米酶材料. 下部分为使用 pubchempy 和 rdkit 文库获得的新描述符扩展了数据库, 表示有机涂层和材料成分的分子特征^[57]; (c) 顶部面板是标准文本挖掘过程的示意图: 左部分是专家注释以构建基线语料库; 中间部分是从文献文本中提取关键信息并构建扩展语料库; 右部分是存储在数据库中以供将来的数据挖掘. 底部面板为将合成句子转换为动作序列的示例. 动作序列的关键组成部分, 如起始和目标材料、合成步骤及其条件, 通过不同的文本挖掘算法从段落中找到和提取^[58]

Fig. 3. (a) Integration of various reaction modules and online analytical tools in a flow chemistry system, forming a continuous synthetic pathway^[54]. (b) The upper part illustrates the DiZyme workflow: from formulating scientific tasks to discovering new nanozyme materials. The lower part shows the expansion of the database with new descriptors obtained using the pubchempy and rdkit libraries, representing molecular features of organic coatings and material compositions^[57]. (c) The top panel is a schematic diagram of the standard text mining process: the left part involves expert annotation to construct a baseline corpus; the middle part extracts key information from literature texts and builds an extended corpus; the right part stores the data in a database for future data mining. The bottom panel provides an example of converting synthesis sentences into action sequences. Key components of the action sequences, such as starting and target materials, synthesis steps, and their conditions, are identified and extracted from paragraphs using different text mining algorithms^[58].

与 DL 的通用材料分类工作流程, 应用于金属玻璃 (metallic glass, MG) 的分类任务中. 研究通过提示工程优化 LLMs 生成的文本描述, 使用 BERT 模型进行分类训练. 结果表明, 在块状金属玻璃 (bulk metallic glass, BMG) 的识别中, 金属模型如图 4(d), 准确率相比传统机器学习模型提高了 463%. 模型预测了 MG 的玻璃形成能力相关性质, 对 101 种经实验验证可形成 BMG 的合金成分进行预测, 发现训练后的 MgBERT 模型能成功地预测全部属于 BMG 类, 有助于对于金属玻璃性质的预测, 模型基本结构如图 4(a). 在电催化领域,

Ding^[63] 于 2023 年提出了一种用于预测电催化剂性质的工作流程 InCrEDible-MaT-GO, 工作流程如图 5(c). 该流程利用人工神经网络模型结合遗传算法进行性质预测.

3.2.3 逆向设计加速材料生成

逆向设计是一种通过已知输出来推断输入或过程的技术. 在材料科学中, 逆向设计被应用于从目标材料的性质出发, 设计和合成具有这些性质的新材料. 近年来, ML 特别是 LLMs 和生成模型的进步, 为逆向设计提供了新的工具和方法.

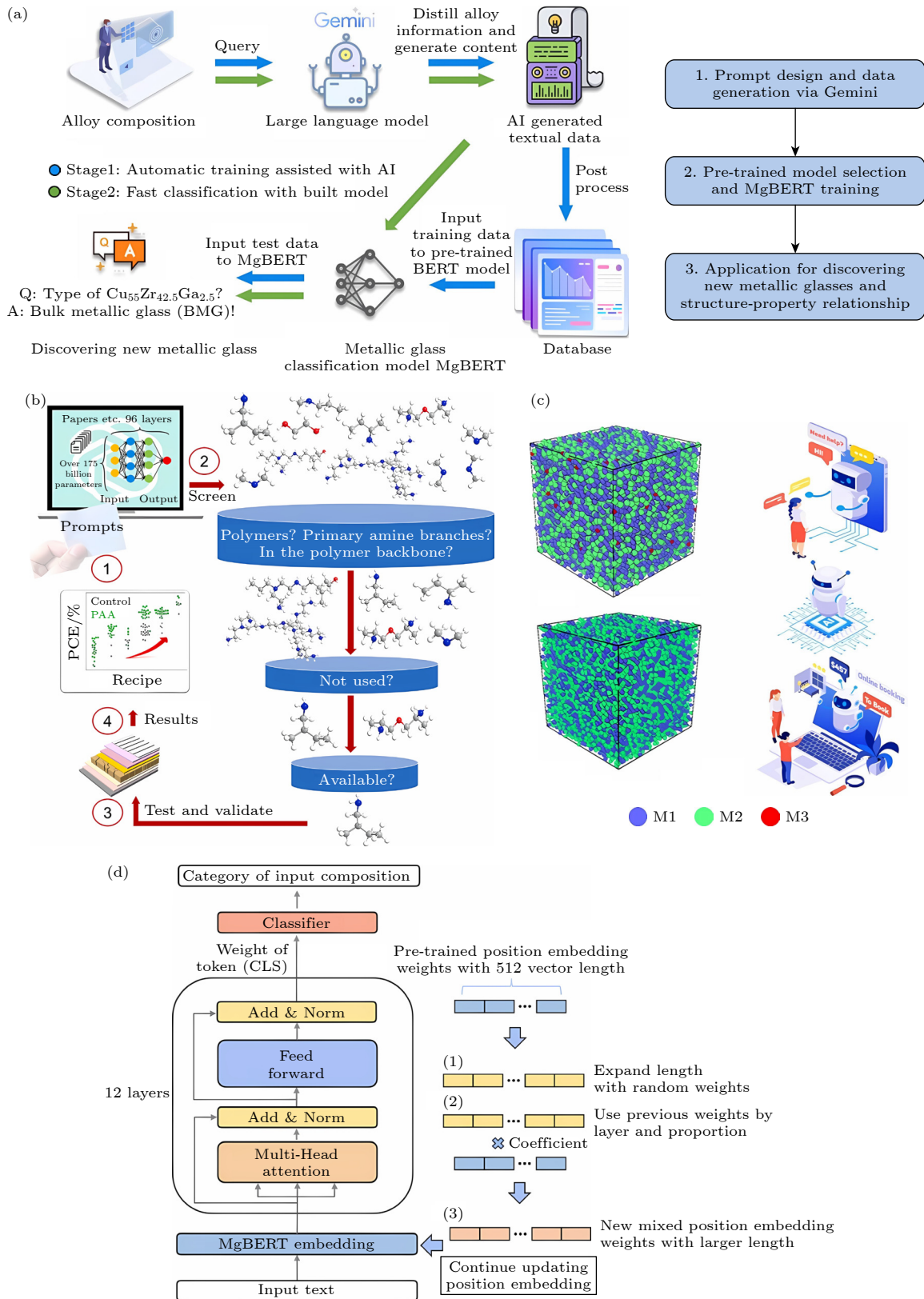


图 4 (a) 生成式 AI 工具可以增强材料科学中的假设生成^[59]; (b) 定制工作流程的示意图, 从已知的合金成分到发现新的金属玻璃^[60]; (c) 玻璃金属模型^[61]; (d) MgBERT 模型的基本架构^[59]

Fig. 4. (a) Generative AI tools can enhance hypothesis generation in materials science^[59]; (b) schematic diagram of a customized workflow, from known alloy compositions to the discovery of new metallic glasses^[60]; (c) a metallic glass model^[61]; (d) the basic architecture of the MgBERT model^[59].

进行逆向设计的第一步是提取信息,并通过信息分析从性质到目标产物的联系,进而进行逆向设计. Ummi 等^[64]利用 AI 助手分析现有的科学文献,并提取关键信息,进而进行逆向设计. 为了解决材料研发中的效率瓶颈、复杂系统建模、数据碎片化、定向设计四大核心难题, Han 等^[65]利用 AI 驱动的材料逆向设计通过整合生成模型、图神经网络与高通量计算,实现了从性质到结构的高效映射,实现了材料科学迈入“按需定制”时代. Kamal^[66]强调了 AtomGPT 在材料设计中的两个关键环节的应用:正向设计(从结构预测性质)和逆向设计(从性质生成结构),其中的逆向设计旨在解决传统材料设计中依赖试错法效率低下的问题,通过结合目标性质(如超导转变温度)直接生成候选原子结构. Bran 等^[36]在文章中介绍的 ChemCrow 是一个展示生成模型在逆向设计中的实际应用的典型例子. ChemCrow 可以根据给定的性质(如溶解度、熔点等)来预测分子结构,并通过实验验证这些预测. 除了 ChemCrow 之外, Oliveira Jr 等^[67]在其文章中提到了如 Stable Diffusion 模型和 DALL-E2 在逆向设计中的应用. 这些模型通过学习大量数据集中的统计分布,能够生成与训练数据相似的新样本. 在材料科学领域,这意味着可以从已知的分子结构出发,生成具有相似或改进性质的新分子结构. 这种方法为研究人员提供了更多的选择和可能性,有助于推动材料科学的创新和发展.

3.2.4 加速计算与建模辅助推动材料科学发展

建模、加速计算和优化辅助技术是推动材料科学快速发展的关键因素^[68]. 近年来,随着人工智能与材料科学的深度融合,诸如多模态学习和 LLMs 等方法在材料计算和优化中展现出显著优势. Cephall 是一个典型的处理多模态视觉语言的平台,用于结构的加速计算和材料结构优化. 在断裂力学和蛋白质力学研究中,能分析微观结构变化图像并阐释关键信息,还可通过图像生成提示可视化新材料设计,加快迭代速度,提供多种方案提高设计准确性和可靠性. 2024 年, Zhang 等^[69]提出的 BERT 模型,其工作流程如图 5(a),通过预训练和微调在虚拟筛选任务上表现优异. 在加速计算和优化辅助方面, Choi 和 Lee^[70]在研究中利用 LLMs 来革新材料语言处理(materials language processing, MLP)

的方法,其流程如图 5(b). 通过精细设计的提示工程,此模型在仅有少量标注数据的情况下,就能实现高效的文本分类、NER 和提取式问答. 在新材料的研发过程中,通过以上技术的应用,从最初的灵感启发到具体的结构设计,再到性能的预测和优化,各个环节都得到了显著的加速.

在材料设计的底层计算优化中, DL 技术通过替代传统量子力学计算流程展现出显著效率优势. 例如, UMM-DFT 模型利用深度神经网络直接预测哈密顿量矩阵,将单点能计算速度提升 3 个数量级,同时保持与 DFT 相当的精度^[71]. 这一突破使得高通量电子结构筛选成为可能,为 LLMs 驱动的合成路径设计提供了可靠的物性数据支撑. 与 DL 模型聚焦数值计算不同, LLMs 的优势在于跨尺度知识融合. 在建模辅助层面, LLMs 通过解析文献与实验数据构建知识图谱(如“掺杂浓度-电导率”关联规则),生成物理约束引导计算参数选择;而 UMM-DFT 等模型则为 LLMs 提供实时物性验证,形成“语义驱动设计→物理精准验证”的闭环优化体系.

3.3 大语言模型驱动的材料研发流程优化

3.3.1 大语言模型助力数据生成

实验数据生成是运用 LLMs 或相关技术创建用于实验的数据集或特定数据,其方法包括基于规则生成、随机生成和从已有数据生成目标数据. Liu 等^[72]利用通用生成式人工智能(general generative artificial intelligence, GAI)在材料科学中的实验数据生成应用, GAI 快速检索的能力极大地节省了数据收集时间. LLMs 能够生成大量与材料科学相关的文本数据,这些数据随后被用于训练和评估 DL 模型. 其优势在于数据不需要手动提取,而是直接从文本数据中进行学习得到. 2024 年, Buehler^[73]从多源数据生成高质量的图像-文本对. 如使用 PyMuPDF 从 PDF 文档中提取图像和对应的文本块,并利用 LLMs 生成详细的图像描述. 为了提高 LLMs 在生成超长序列数据时的效率, Wu 等^[74]在 2025 年提出了 TOKENSWIFT 框架,通过多 token 生成、动态键值缓存管理和上下文惩罚等技术手段,显著加速了 LLMs 生成超长序列的过程,从而提升了 LLMs 在数据生成任务中的性能,使得其能够更高效地生成长篇文本数据.

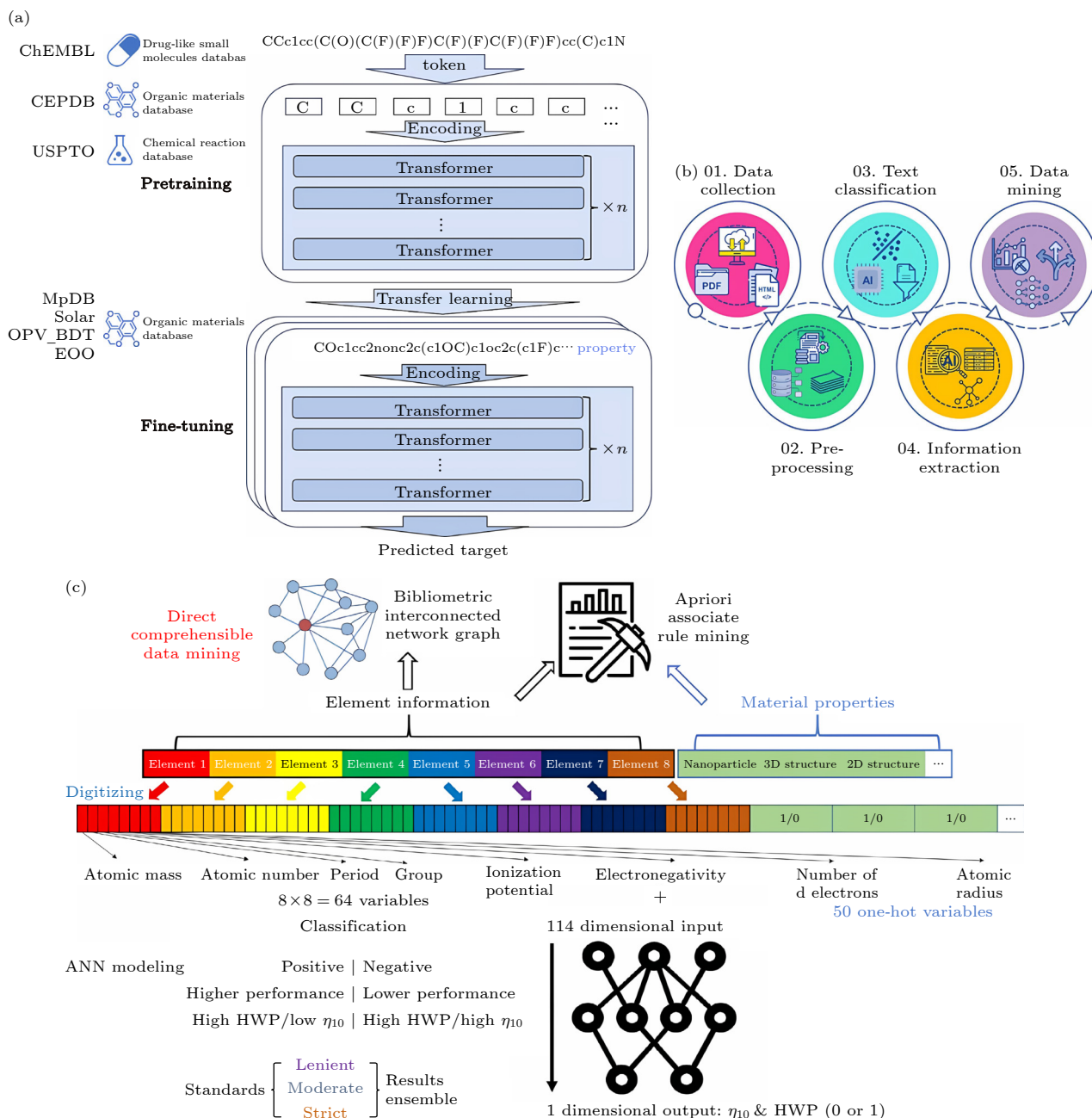


图 5 (a) BERT 的工作流程^[69]; (b) MLP 的过程包括五个步骤: 数据收集、预处理、文本分类、信息提取和数据挖掘^[70]; (c) 用于构建 ANN 建模数据库的数字化过程图示^[63]

Fig. 5. (a) Workflow of BERT^[69]; (b) the process of MLP consists of five steps: data collection, preprocessing, text classification, information extraction, and data mining^[70]; (c) illustration of the digitization process for constructing an ANN modeling database^[63].

3.3.2 大语言模型驱动的实验设计

优化实验设计是利用 LLMs 等技术, 通过生成实验参数和条件实现流程自动化, 挖掘知识辅助决策的方式, 对材料实验流程进行改进完善以提升效率和质量的管理活动. 2024 年, Kiichi 等^[75]介绍了一种材料开发平台, 该平台能够在数据有限的情况下进行准确的材料属性预测, 从而优化实验设计. LLMs 的推理能力和代理的自主性能能够在实验

设计阶段提供指导, 优化参数选择、减少实验成本、加速新材料的发现. 为了解决研究复杂性问题、突破人工编码效率瓶颈, Li 等^[76]提出了一个基于 LLMs 的端到端自动化机器学习研究框架: MLR-Copilot, 该框架首先由 IdeaAgent 分析文献生成创新假设与实验计划, 再通过 Experiment-Agent 动态检索原型代码、模型及数据集, 最后在反馈驱动的执行阶段实现实验迭代优化与人机协

同验证. 该框架在语义分析、图像识别等 5 类任务中验证有效性, 实验显示假设创新性评分提升 26%(3.9 vs. 3.1), 情感分析任务准确率较基线提升 78.5%, 首次实现从假设生成到结果验证的全链路自动化, 为研究者提供高效可靠的“AI 协作者”范式. Luo 等^[7]通过 LLMs 中的模块化和工具集成, 优化实验设计并自动化实验流程, 显著提升了科学研究的效率. 这些技术的结合不仅提高了实验的效率和准确性, 还促进了数据驱动的科学研究, 推动了催化剂设计和优化的革命性变革.

4 结论与展望

LLMs 在材料设计领域的应用正引发研究范式的深刻变革, 展现出推动材料科学发展、革新传统设计流程的巨大潜力. 通过信息检索与知识理解技术, LLMs 能够从海量文本数据中高效提取关键信息, 助力材料知识的精准获取与深度整合. 在智能问答、信息提取环节, 基于知识图谱等技术, 模型为研究人员提供了便捷、智能的知识交互途径, 有效解决材料设计中的各类问题, 拓宽研究思路. 知识图谱的构建与维护进一步强化了知识管理, 为材料设计构建起全面、系统的知识体系, 确保知识的时效性与准确性.

在材料设计的核心环节, 如合成预测、结构与性质预测, LLMs 凭借对大量数据的学习与分析, 突破传统方法的局限, 为材料合成路径规划、微观结构解析以及性能优化提供了创新性解决方案. 通过优化实验设计、自动化数据生成以及数据格式转换等工作流程优化手段, LLMs 实现了与实验流程的深度融合, 显著提升了材料研发的效率与质量, 加速了从理论设计到实际制备的进程. LLMs 在智能化标注数据集方面的应用, 为高质量数据的获取提供了新途径, 解决了传统人工标注的痛点, 保障了模型训练数据的可靠性. 在智能实验平台构建中, 结合 NLP 技术, LLMs 实现了与人、实验机器人及智能平台的高效交互, 推动材料研发向自动化、智能化迈进.

在实际应用中, LLMs 通过跨尺度知识融合与智能推理, 挖掘数据中隐含的规律与模式, 凭借在加速计算与合成设计方面的优势, 驱动材料性质预测与逆向设计实现关键技术突破. 例如, 在合成预测与路线设计方面, LLMs 可借助文本挖掘技术和

先进的机器算法从海量的科学文献中提取材料合成所需的数据, 据此可对目标材料的合成路径与关键条件进行逆向设计与推算, 简化材料发现流程. 在结构与性质预测方面, LLMs 在结构预测中的潜力已经在材料科学的应用中得以展现, 通过将生成的文本与现有的科学文献相结合, 研究人员能高效筛选候选材料结构, 并提出可经实验验证的假设. 在性质预测方面, LLMs 能够结合结构信息和物理化学原理, 通过分析材料的晶体结构和电子态预测其物理、化学性质, 科学家可以从庞大的实验数据集中构建出精确的预测模型, 通过训练模型来高效学习结构与性质之间的关系.

然而, 当前研究仍存在一些挑战. 一方面, LLMs 对高质量数据的依赖较为严重, 数据的准确性、完整性以及多样性不足制约了模型性能的进一步提升, 尽管面临计算资源需求高和可解释性差等挑战, 但随着技术进步, LLMs 与 NER 的结合有望在材料科学领域发挥更大作用. 另一方面, 模型的可解释性问题亟待解决, 其内部决策机制的“黑箱”特性阻碍了与专业知识的深度融合, 不利于研究人员深入理解和应用模型. 此外, LLMs 在处理复杂材料体系的深层次理解和精准设计时, 仍显得力不从心, 模型性能的进一步提升受到数据质量与数据规模不足的制约.

展望未来, 随着技术的不断发展, LLMs 有望在材料设计领域取得更大突破. 在数据层面, 应加强多源数据融合与清洗, 构建标准化、大规模的高质量材料数据库, 为模型提供坚实的数据支撑. 在算法优化上, 结合材料科学原理, 对模型架构进行针对性改进, 引入物理、化学等先验知识, 提高模型预测的科学性与合理性. 同时, 加大在可解释性研究方面的投入, 开发可视化工具与解释性算法, 使模型决策过程透明化, 促进模型与专业知识的有机结合. 此外, 应进一步推动 LLMs 与自动化实验平台的深度融合, 实现从实验设计到数据采集、分析的全流程智能化, 提高材料研发效率. 在跨学科合作方面, 加强材料科学与计算机科学、物理学、化学等多学科的交叉合作, 拓展 LLMs 在更多材料领域和应用场景中的应用, 为材料创新和应用开辟更广阔的前景. 另外, LLMs 需进一步与物理约束驱动的机器学习模型深度融合, 以提升材料设计的科学合理性. 如等变神经网络因其对原子系统对称性的严格保持, 可作为 LLMs 生成结构的物理校验

模块, 避免生成违反基本物理规律的材料构型^[78]. 通过将等变神经网络的归纳偏置引入 LLMs 的生成过程, 有望实现“数据驱动创新”与“第一性原理约束”的平衡, 推动可解释性材料设计范式的发展.

综上所述, LLMs 为材料设计带来了前所未有的机遇, 尽管面临挑战, 但通过持续的技术创新与跨学科合作, 有望在未来彻底改变材料研发的模式, 为满足各领域对高性能材料的迫切需求提供强有力的技术保障, 推动材料科学迈向全新的发展阶段.

参考文献

- [1] Lewis P, Perez E, Piktus A, Petroni F, Karpukhin V, Goyal N, Kuttler H, Lewis M, Yin W, Rocktaschel T, Riedel S, Kiela D 2020 *34th Conference on Neural Information Processing Systems (NeurIPS) Canada*, December 6–12, 2020, p16792
- [2] Shi L, Liu Z M, Yang Y, Wu W Z, Zhang Y Y, Zhang H B, Lin J, Wu S Y, Chen Z H, Li R M, Wang N, Liu Z P, Tan H B, Gao H Y, Zhang Y, Wang G 2024 arXiv: 2408.04665[cs.CL]
- [3] Luu R K, Buehler M J 2024 *Adv. Sci.* **11** e2306724
- [4] Li J T, Liu Y Q, Fan W Q, Wei X Y, Liu H, Tang J L, Li Q 2024 *IEEE Trans. Knowl. Data Eng.* **36** 6071
- [5] Park N H, Callahan T J, Hedrick J L, Erdmann T, Capponi S 2024 arXiv: 2408.11793[cs. AI]
- [6] Chiang Y, Hsieh E, Chou C H, Riebesell J 2024 arXiv: 2401.17244[cs. CL]
- [7] Tang Y H, Xu W B, Cao J, Ma J Z, Gao W L, Farrell S, Erichson B, Mahoney M W, Nonaka A, Yao Z 2025 arXiv: 2502.13107[cs. AI]
- [8] Li C T, Han X, Jiang R H, Yun P W, Hu P F, Ban X J 2023 *Chin. J. Eng. Sci.* **46** 290 (in Chinese) [李长泰, 韩旭, 蒋若辉, 贡培文, 胡鹏飞, 班晓娟 2023 *工程科学学报* **46** 290]
- [9] Tanishq G, Mohd Z, Anoop Krishnan N M, Mausam 2022 *npj Comput. Mater.* **8** 940
- [10] Lai N S, Tew Y S, Zhong X L, Yin J, Li J L, Yan B H, Wang X N 2023 *Ind. Eng. Chem. Res.* **62** 17835
- [11] Thway M, Low A K Y, Khetan S, Dai H W, Recatala J, Chen A P, Hippalgaonkar K 2024 *Digital Discovery* **3** 328
- [12] Yu S L, Ran N, Liu J J 2024 *Artif. Intell. Chem.* **2** 100076
- [13] Durmaz A R, Thomas A, Mishra L, Murthy R N, Straub T 2024 *Sci. Data.* **11** 1112
- [14] Lei G, Docherty R, Cooper S J 2024 *Digital Discovery* **3** 1257
- [15] Weston L, Tshitoyan V, Dagdelen J, Kononova O, Trewartha A, Persson K A, Ceder G 2019 *J. Chem. Inf. Model.* **59** 3692
- [16] Zeng Z N, Yin B C, Wang S P, Liu J R, Yang C, Yao H S, Sun X Z, Sun M S, Xie G T, Liu Z Y 2024 *Bioinformatics* **40** btac534
- [17] Jia X, Aziz A, Hashimoto Y, Li H 2024 *Sci. China Mater.* **67** 1173
- [18] Polak M P, Morgan D 2024 *Nat. Commun.* **15** 1569
- [19] Foppiano L, Lambard G, Amagasa T, Ishii M 2024 *Sci. Technol. Adv. Mater. Methods.* **4** 2356506
- [20] Dagdelen J, Dunn A, Lee S, Walker N, Rosen A S, Ceder G, Persson K A, Jain A 2024 *Nat. Commun.* **15** 1418
- [21] Zheng Z L, Zhang O F, Borgs C, Chayes J T, Yaghi O M 2023 *J. Am. Chem. Soc.* **145** 18048
- [22] Yang J M, Walker K C, Bekar A A, Hao B, Bhadelia N, Joseph D, Paschalidis I C, 2024 *Int. J. Med. Informatics* **189** 105500
- [23] Shi Z B, Zhu L Y, Le X Q 2024 *Data Anal. Knowl. Discovery* **8** 23 (in Chinese) [时宗彬, 朱丽雅, 乐小虬 2024 *数据分析与知识发现* **8** 23]
- [24] Buehler M J 2023 *J. Mech. Phys. Solids* **181** 105454
- [25] Zia G A J, Valdestilhas A, Torres B J M, Kruschwitz S 2024 *The 1st International Workshop on Semantic Materials Science co-located with the 20th International Conference on Semantic Systems (SEMANTiCS) Amsterdam, The Netherlands, September 17–19, 2024* p101
- [26] Corlatescu D G, Watanabe M, Ruseti S, Dascalu M, McNamara D S 2024 *Comput. Hum. Behav.* **154** 108154
- [27] Bai X F, He S, Li Y, Xie Y B, Zhang X, Du W L, Li J R 2025 *NPJ Comput. Mater.* **11** 1
- [28] He Q, Yang X Q, Xu Y 2018 *Fiber Reinf. Plast./Compos.* **62** (in Chinese) [贺强, 杨晓强, 徐艺 2018 *玻璃钢/复合材料* **62**]
- [29] Chen H L 2017 *M. S. Thesis (Shijiazhuang: Hebei University of Science and Technology)* (in Chinese) [陈慧琳 2017 硕士学位论文 (石家庄: 河北科技大学)]
- [30] Ye Y P, Ren J, Wang S Z, Wan Y W, Wang H F, Razzak I, Hoex B, Xie T, Zhang W J 2024 arXiv: 2404.03080[cs. CL]
- [31] Yang F L, Egon C, Xue J, Ryuhei S, Kazuaki K, Yusuke H, Shin O, Li H 2023 *Nano Mater. Sci.* **6** 256
- [32] Songshan Lake Materials Laboratory, Chinese Academy of Sciences Institute of Physics <https://news.qq.com/rain/a/20250211A03D5O00> [2025-2-11]
- [33] Buehler M J 2024 *ACS Eng. Au* **4** 241
- [34] Cai J M, Yuan Y J, Sui X P, Lin Y Z, Zhuang K, Xu Y, Zhang Q, Ukrainczyk N, Xie T Y 2024 *Constr. Build. Mater.* **425** 135965
- [35] Ren H Y, Liu J P, Wang J, Gu X X, Chen X, Zhang Y, Zhao C X 2025 *Comput. Eng. Appl.* **61** 1 (in Chinese) [任海玉, 刘建平, 王健, 顾勋勋, 陈曦, 张越, 赵昌珏 2025 *计算机工程与应用* **61** 1]
- [36] Bran A M, Cox S, Schilter O, Baldassari C, White A D, Schwaller P 2024 *Nat. Mach. Intell.* **6** 525
- [37] Ansari M, Watchorn J, Brown C E, Brown J S 2024 arXiv: 2410.03963v1[physics. chem-ph]
- [38] Zheng Z L, Rampal N, Inizan T J, Borgs C, Chayes J T, Yaghi O M 2025 *Nat. Rev. Mater.* **10** 369
- [39] Yoshikawa N, Skreta M, Darvish K, Arellano S, Zhi J, Kristensen L B, Li A Z, Zhao Y C, Xu H P, Kuramshin A, Aspuru A, Shkurti F, Garg A 2023 *Auton. Robot.* **47** 1057
- [40] Zhu Q, Zhang F, Huang Y, et al. 2022 *Natl. Sci. Rev.* **9** nwac190
- [41] Dai T W, Vijaykrishnan S, Szczypiński F T, Ayme J F, Simaei E, Fellowes T, Clowes R, Kotopanov L, Shields C E, Zhou Z X, Ward J W, Cooper A I 2024 *Nature* **635** 890
- [42] Seifrid M, Strieth F, Haddadnia M, Wu T C, Alca E, Bodo L, Arellano S, Yoshikawa N, Skreta M, Keunen R, Aspuru A 2024 *Digital Discovery* **3** 1319
- [43] Zheng Z L, Florit F, Jin B, Wu Hao Y, Li S C, Nandiwale K Y, Salazar C S, Mustakis J G, Green W H, Jensen K F 2024 *Angew. Chem. Int. Ed.* **64** e202418074
- [44] Ruan Y X, Lu C Y, Xu N, He Y C, Chen Y X, Zhang J, Xuan J, Pan J Z, Fang Q, Gao H Y, Shen X D, Ye N, Zhang Q, Mo Y M 2024 *Nat. Commun.* **15** 10160
- [45] Hatakeyama K, Ishikawa H, Takaishi S, Igarashi Y, Nabae Y, Hayakawa T 2024 *Polym. J.* **56** 997
- [46] Zhou J Y, Luo M, Chen L J, Zhu Q, Jiang S, Zhang F, Shang

- W W, Jiang J 2025 *Digital Discovery* **4** 636
- [47] Antunes L M, Butler K T, Grau C R 2024 *Nat. Commun.* **15** 10570
- [48] Szymanski N J, Rendy B, Fei Y X, et al. 2023 *Nature* **624** 86
- [49] Sriram A, Miller B K, Chen R T Q, Wood B M 2024 arXiv: 2410.23405[cs.LG]
- [50] Jia S Y, Zhang C, Fung V 2024 arXiv: 2406.13163[cond-mat.mtrl-sci]
- [51] Liu H, Zheng H, Jia Z H, Zhou B H, Liu Y, Chen X L, Feng Y J, Li W, Yang W J, Li H 2023 *Front. Chem. Sci. Eng.* **17** 2156
- [52] Zhang C Y, Wang X Y, Wang Z Y 2024 *Chin. J. Catal.* **59** 7 (in Chinese) [张成翼, 王兴宇, 王子运 2024 *催化学报* **59** 7]
- [53] Slautin B N, Liu Y T, Liu Y, Emery R, Hong S, Dubey A, Shvartsman V V, Lupascu D C, Sanchez S L, Ahmadi M, Kim Y, Strelcov E, Brown K A, Rack P D, Kalinin S V 2025 arXiv: 2501.02503[cond-mat.mtrl-sci]
- [54] Su Y M, Wang X, Ye Y X, Xie Y B, Xu Y J, Jiang Y B, Wang C 2024 *Chem. Sci.* **15** 12200
- [55] Chen Z Y, Xie F K, Wan M, Yuan Y, Liu M, Wang Z G, Meng S, Wang Y L 2023 *Chin. Phys. B* **32** 173 (in Chinese) [陈子逸, 谢帆恺, 万萌, 袁扬, 刘森, 王宗国, 孟胜, 王彦桐 2023 *中国物理 B* **32** 173]
- [56] Zhang D, Li H 2024 ChemRxiv. DOI: 10.26434/chemrxiv-2024-9lpb9
- [57] Razlivina J, Dmitrenko A, Vinogradov V 2024 *J. Phys. Chem.* **15** 5804
- [58] Chen X Q, Gao Y, Wang L D, Cui W J, Huang J M, Du Y, Wang B 2024 *Sci. Data* **11** 347
- [59] Liu S Y, Wen T Q, Pattamatta A S L S, Srolovitz D J 2024 *Mater. Today* **80** 240
- [60] Chen C, Maqsood A, Zhang Z, Wang X B, Duan L R, Wang H H, Chen T Y, Liu S Y, Li Q T, Luo J S, Jacobsson T J 2024 *Cell Rep. Phys. Sci.* **5** 102058
- [61] Wang J Y, Liu X J, Wu Y, Wang H, Ma D, Lu Z P 2023 *Acta Mater.* **261** 119386
- [62] Wang Z W, Han M, Jin B 2024 *Environ. Chem.* **43** 69 (in Chinese) [王紫维, 韩民, 金彪 2024 *环境化学* **43** 69]
- [63] Ding R, Wang X B, Tan A, Li J, Liu J G 2023 *ACS Catal.* **13** 13267
- [64] Unni R, Zhou M Y, Wiecha P R, Zheng Y B 2024 *Curr. Opin. Solid State Mater. Sci.* **30** 101157
- [65] Han X Q, Wang X D, Xu M Y, Feng Z, Yao B W, Guo P J, Gao Z F, Lu Z Y 2025 *Chin. Phys. Lett.* **42** 027403
- [66] Choudhary K 2024 *J. Am. Chem. Soc.* **15** 6909
- [67] Oliveira O N Jr, Christino L, Oliveira M C F, Paulovich F V 2023 *J. Chem. Inf. Model.* **63** 7605
- [68] Liu G, Sun M, Wojciech M, Jiang M, Chen J 2024 arXiv: 2410.04223 [cs.LG]
- [69] Zhang C W, Zhai Y S, Gong Z Y, Duan H L, She Y B, Yang Y F, Su A 2024 *J. Cheminformatics* **16** 89
- [70] Choi J, Lee B 2024 *Commun. Mater.* **5** 13
- [71] Wang Y X, Li Y, Tang Z C, Li H, Yuan Z L, Tao H G, Zou N L, Bao T, Liang X H, Chen Z Z, Xu S H, Bian C, Xu Z M, Wang C, Si C, Duan W H, Xu Y 2024 *Sci. Bull.* **69** 2514
- [72] Liu Y, Yang Z W, Yu Z Y, Liu Z T, Liu D H, Lin H L, Li M Q, Ma S C, Avdeev X, Shi S Q 2023 *J. Materiomics* **9** 798
- [73] Buehler M J 2024 *Adv. Funct. Mater.* **34** 9531
- [74] Wu T, Shen J Z, Jia Z X, Wang Y X, Zheng Z L 2025 arXiv: 2502.18890[cs.CL]
- [75] Obuchi K, Funaya K, Toyama K 2024 *NEC Tech. J.* **17** 46
- [76] Li R C, Patel T, Wang Q Y, Du X Y 2024 arXiv: 2408.14033 [cs.AI]
- [77] Luo Z M, Yang Z L, Xu Z X, Yang W, Du X Y 2025 arXiv: 2501.04306[cs.CL]
- [78] Tang Z C, Li H, Lin P Z, Gong X X, Jin G, He L X, Jiang H, Ren X G, Duan W H, Xu Y 2024 *Nat. Commun.* **15** 8815

SPECIAL TOPIC—AI + Physical Science

Material design accelerated by large language models: end-to-end empowerment from knowledge mining to intelligent design^{*}

HUANG Yudan¹⁾ XIA Wanjun²⁾ DU Junmei¹⁾ JIANG Yu¹⁾ WANG Xin¹⁾
CHEN Yuanzheng¹⁾ WANG Hongyan¹⁾ ZHAO Jijun^{3)†} GUO Chunsheng^{1)2)‡}

1) (*School of Physical Science and Technology, Southwest Jiaotong University, Chengdu 611756, China*)

2) (*Library of Southwest Jiaotong University, Chengdu 611756, China*)

3) (*School of Physics, South China Normal University, Guangzhou 510631, China*)

(Received 17 April 2025; revised manuscript received 3 June 2025)

Abstract

With the rapid development of artificial intelligence technology, large language models (LLMs) have become the core driving force for the paradigm shift in materials science research. This review explores the comprehensive role of LLMs in accelerating material design throughout the entire research lifecycle from knowledge mining to intelligent design. This work aims to emphasize how LLMs can leverage their advantages in information retrieval, cross-modal data integration, and intelligent reasoning to address challenges in traditional materials research, such as data fragmentation, high experimental costs, and limited reasoning capabilities.

Key methods include applying LLMs to knowledge discovery through techniques such as retrieval-augmented generation (RAG), multi-modal information retrieval, and knowledge graph construction. These approaches can efficiently extract and construct material data from a vast repository of scientific literature and experimental records. Additionally, LLMs are integrated with automated experimental platforms to optimize workflows from natural language-driven experiment design to high-throughput iterative testing.

The results demonstrate that LLMs significantly enhance material research efficiency and accuracy. For instance, in knowledge mining, LLMs improve information retrieval accuracy by up to 29.4% in tasks such as predicting material synthesis conditions. In material design, LLMs can accelerate computational modeling, structure and performance prediction, and reverse engineering, reducing experimental trial-and-error cycles. Notably, LLMs perform well in cross-scale knowledge integration, linking material composition, processing parameters, and performance metrics to guide innovative synthesis pathways.

However, challenges still exist, including dependence on high-quality data, the “black-box” nature of LLMs, and limitations in handling complex material systems. The future direction emphasizes improving data quality through multi-source integration, enhancing model explainability through visualization tools, and deepening interdisciplinary collaboration, and bridging the gaps between AI and domain-specific expertise.

In summary, LLMs are reshaping materials science by implementing a data-driven, knowledge-intensive research paradigms. The ability of LLMs to integrate vast datasets, predict material properties, and automate experimental workflows makes them indispensable tools for accelerating material discovery and innovation. With the development of LLMs, their synergistic effect with physical constraints and experimental platforms is expected to open new fields in material design.

Keywords: large language models, material design, intelligent tasks, synthesis and property prediction

PACS: 81.05.Zx, 81.20.-n

DOI: [10.7498/aps.74.20250497](https://doi.org/10.7498/aps.74.20250497)

CSTR: [32037.14.aps.74.20250497](https://cstr.cn/32037.14.aps.74.20250497)

^{*} Project supported by the National Natural Science Foundation of China (Grant No. U2167217) and the Sichuan Academic Achievement Analysis and Application Research Center, China (Grant No. SCAA24-B19).

[†] Corresponding author. E-mail: zhaojj@scnu.edu.cn

[‡] Corresponding author. E-mail: csguo@swjtu.edu.cn



大语言模型加速材料设计——从知识挖掘到智能设计的全链条赋能

黄钰丹 夏琬钧 杜俊梅 蒋渝 汪鑫 陈元正 王红艳 赵纪军 郭春生

Material design accelerated by large language models: end-to-end empowerment from knowledge mining to intelligent design

HUANG Yudan XIA Wanjun DU Junmei JIANG Yu WANG Xin CHEN Yuanzheng WANG Hongyan
ZHAO Jijun GUO Chunsheng

引用信息 Citation: *Acta Physica Sinica*, 74, 188101 (2025) DOI: 10.7498/aps.74.20250497

CSTR: 32037.14.aps.74.20250497

在线阅读 View online: <https://doi.org/10.7498/aps.74.20250497>

当期内容 View table of contents: <http://wulixb.iphy.ac.cn>

您可能感兴趣的其他文章

Articles you may be interested in

基于FTO/Ag/FTO构型的高透明红外隐身薄膜设计

Design of high transparent infrared stealth thin films based on FTO/Ag/FTO structure

物理学报. 2023, 72(24): 244202 <https://doi.org/10.7498/aps.72.20231084>

大语言模型在电池科研全流程应用的测评与无机固态电解质综合数据库构建

Evaluation of the application of large language models in the entire process of battery research and development of a comprehensive database for inorganic solid electrolyte

物理学报. 2025, 74(16): 160701 <https://doi.org/10.7498/aps.74.20250572>

大尺寸三维超声振动系统的智能优化设计

Intelligent optimization design of large-scale three-dimensional ultrasonic vibration system

物理学报. 2024, 73(8): 084304 <https://doi.org/10.7498/aps.73.20240006>

基于移动渐近线算法的大角度偏转分束器的智能设计

Intelligent design of large angle deflection beam splitter based on method of moving asymptotes

物理学报. 2021, 70(23): 234102 <https://doi.org/10.7498/aps.70.20210117>

片上光互连器件的智能化设计研究进展

Research progress of intelligent design of on-chip optical interconnection devices

物理学报. 2023, 72(18): 184204 <https://doi.org/10.7498/aps.72.20230705>

融合结构知识的蛋白质预训练模型进展

Progress in protein pre-training models integrating structural knowledge

物理学报. 2024, 73(18): 188701 <https://doi.org/10.7498/aps.73.20240811>