

特色专题

数据驱动的材料研发新范式:高通量计算与机器学习融合创新

萨百晟¹, 陈林涛¹, 祝令刚², 杨明理³, 宿彦京⁴, 周健², 孙志梅^{2*}, 谢建新⁴

摘要 人工智能技术的快速发展正在推动新一轮科技革命,并促进以材料基因工程为代表的驱动材料研发体系由经验驱动模式向数据驱动与智能设计协同推进模式转变。围绕高通量计算与机器学习融合驱动下的材料研发范式变革,系统梳理了材料研发范式的演进过程、高通量计算与机器学习的关键技术体系、国内外战略布局与平台建设进展,并分析其对新材料产业生态、科研组织模式及重点材料研发路径的影响。高通量计算支撑材料数据生成与大规模筛选,机器学习则通过快速预筛选和知识挖掘促进材料智能设计闭环迭代,推动材料创新竞争由单一技术能力竞争转向数据资源、计算工具、研发平台与组织体系协同能力的综合竞争。中国已初步具备相关建设基础与竞争力,后续可围绕核心软件自主化、高质量数据积累、标准体系完善,以及研发链与产业链贯通等方面持续提升。高通量计算与机器学习的深度融合,正在推动材料研发由第四范式向第五范式演进,并持续重塑新材料创新的技术体系、产业生态与科研组织逻辑。

关键词 材料基因工程;高通量计算;机器学习;材料研发范式;材料数据库

习近平总书记指出,新材料产业是战略性、基础性产业,也是高技术竞争的关键领域^[1]。当前,数字技术和人工智能(artificial intelligence, AI)与科学研究深度融合,正在推动材料研发范式变革。传统材料研发主要遵循“经验试错—实验验证—性能优化”的串行路径,面对高维成分空间、多尺度结构和复杂工艺时,逐渐暴露出研发周期长、试错成本高、知识分散和跨体系迁移困难等问题。随着高性能计算、材料数据库

和 AI 技术的发展,材料研发正在转向数据驱动与智能设计协同推进的新范式^[2-3]。

材料基因工程的核心是将高通量计算、高通量实验和材料数据库相结合,通过系统积累并关联材料成分、结构、工艺与性能数据,提高候选材料筛选和优化效率^[4]。随着 AI 发展,由此形成的数据资源、自动化流程和平台体系进一步成为材料智能设计和人机协同研发的基础,推动材料科学由数据密集型研究迈向智能化科研(AI for Research, AI4R)。

在这一演进过程中,高通量计算与机器学习的融合构成材料研发新范式的关键技术主线。高通量计算通过多尺度模拟和自动化 workflow 生成材料数据,机器学习从多源异构数据中提取结构—性能关联,并前

1. 福州大学材料科学与工程学院,福州 350108

2. 北京航空航天大学材料科学与工程学院,北京 100083

3. 四川大学材料基因工程研究中心,成都 610065

4. 北京科技大学高精尖学院,北京 100083

收稿日期: 2026-05-11; 修回日期: 2026-07-06

基金项目: 国家科技重大专项(2025ZD0618802); 中国工程院、国家自然科学基金委联合战略研究咨询项目(2025-XZ-33)

作者简介: 萨百晟,教授,研究方向为材料高通量计算与人工智能,电子信箱: bssa@fzu.edu.cn; 孙志梅(通信作者),教授,研究方向为先进材料的人工智能,电子信箱: zmsun@buaa.edu.cn

引用格式: 萨百晟,陈林涛,祝令刚,等. 数据驱动的材料研发新范式: 高通量计算与机器学习融合创新[J]. 科技导报, 2026, 44(14): 91-103;

doi:10.3981/j.issn.1000-7857.2026.05.00035

移材料设计和性能判断环节。材料研发不再局限于单点计算或经验筛选,二者结合形成“计算设计—智能预测—实验验证—反馈优化”的闭环,并为第五范式下 AI 参与问题分解、候选生成和规律发现提供支撑。

围绕这一趋势,世界主要经济体已将高通量计算、材料数据基础设施和 AI 纳入材料创新战略。中国自“十三五”材料基因工程重大专项以来,持续推进材料数字化研发、平台建设和 AI 应用,初步形成材料智能研发的政策与技术基础^[5-6]。本文围绕高通量计算与机器学习融合,梳理研究范式、关键技术、国际竞争、平台建设及产业影响,并从数据资源、计算工具、机器学习模型和产业应用协同等角度,分析中国构建自主可控材料智能研发体系的基础、挑战与发展路径。

1 研究范式的演进

1.1 材料研发范式的历史演进

材料科学经历了由“经验试错”向“数据驱动”再向“AI4R”演进的过程(图 1)。前 3 种范式分别依托实验经验、基础理论和计算模拟推动材料性能优化。第四范式兴起于 21 世纪初,以美国材料基因组计划(Materials Genome Initiative, MGI)的启动为重要标志,依托“高通量计算—高通量实验—材料数据库”三位一体架构,借助大数据和机器学习挖掘材料成分—结构—性能关联^[2,4]。材料基因工程由此成为第四范式在材料科学中的集中体现,其核心是通过计算、实验与数据协同提升材料发现效率。

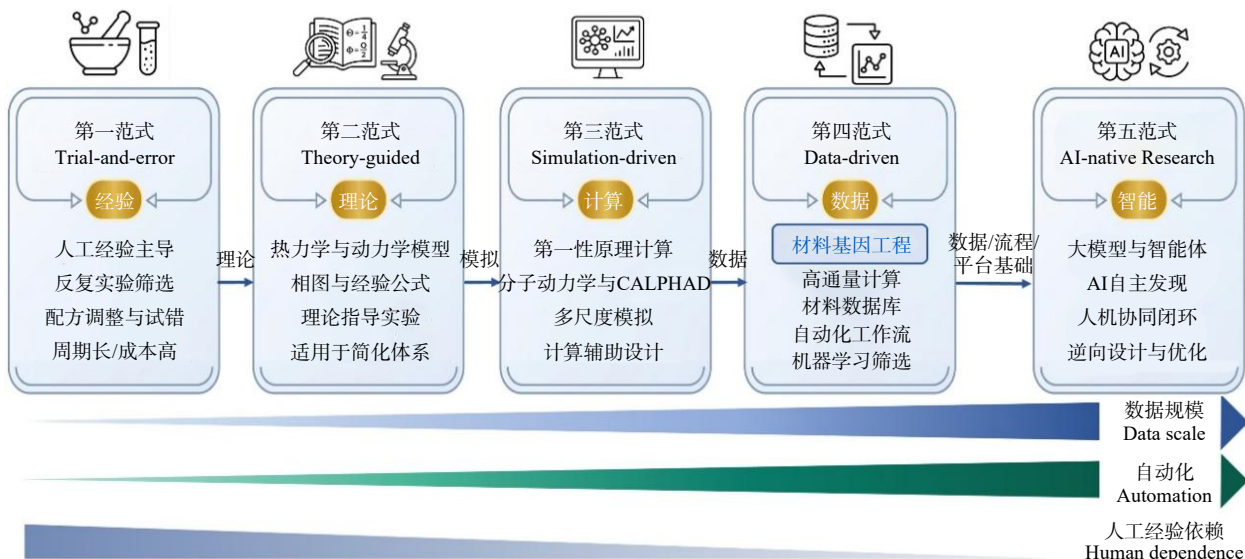


图 1 材料研发范式的演进过程

当前,材料研发正在迈向第五范式,即 AI4R 阶段。李国杰^[7]将其概括为 AI 全面融入科研、人机智能融合、应对组合爆炸、跨学科协同和大模型平台支撑;鄂维南^[8]和张平文^[9]等强调, AI for Science 的关键并非简单替代传统计算或实验,而在于与具体科学问题、领域知识、计算平台和机理模型深度结合。第五范式与第四范式的差异不仅在于算法性能或计算规模提升,还体现在材料知识生成方式变化。第四范式主要依托高通量计算、高通量实验和材料数据库,在人工设定目标、描述符和评价规则下开展大规模筛选, AI 更多承担加速计算、模式识别和候选预筛选;第五范式则推动 AI 进一步进入问题分解、规律发

现、候选生成和假设提出等环节,由辅助工具转向科研过程的重要参与者^[8,10]。

在材料研发中,这种转变具体体现为“模型—计算—实验—知识”的闭环重构。材料大模型和领域知识模型可从文献、数据库和实验记录中提取组成、结构、工艺与性能关系,为候选生成和研究方案设计提供知识基础;机器学习势函数和图神经网络可加速原子尺度稳定性、动力学过程和关键性质预测;主动学习则根据模型不确定性选择最具信息增益的计算或实验样本,推动高通量计算、高通量实验和模型训练之间形成迭代反馈^[11]。图网络材料探索模型(graph networks for materials exploration, GNoME)表明, AI

的作用正由性能预测延伸至候选生成和稳定性筛选^[10,12]。

AI 在材料研发链条中的作用不断前移,其可靠应用也对数据质量、物理约束和实验验证提出更高要求,但第五范式并不意味着其能脱离材料科学知识独立完成可靠发现。材料体系普遍存在缺陷、界面、非平衡相、工艺历史和服役环境耦合等复杂因素,AI 生成的候选材料和研究假设仍需满足物理可解释性、实验可验证性和可重复性要求。模型适用域、黑箱模型解释边界、分布外泛化能力、统计相关性向因果机制

转化的可靠性,以及数据产权、共享边界、模型偏差、自动化实验责任和高风险材料设计审查,共同界定 AI 深度参与材料研发的实际边界。

1.2 材料高通量计算技术体系

材料高通量计算为第四与第五范式提供数据支撑,其核心是借助自动化流程和高性能算力,将单点计算扩展为面向材料成分、结构与性能的大规模并行筛选。它由高通量建模、多尺度计算、自动化 workflow、数据管理、结果分析、智能纠错和机器学习模型共同支撑,形成材料数据生成与筛选体系(图 2)^[13]。

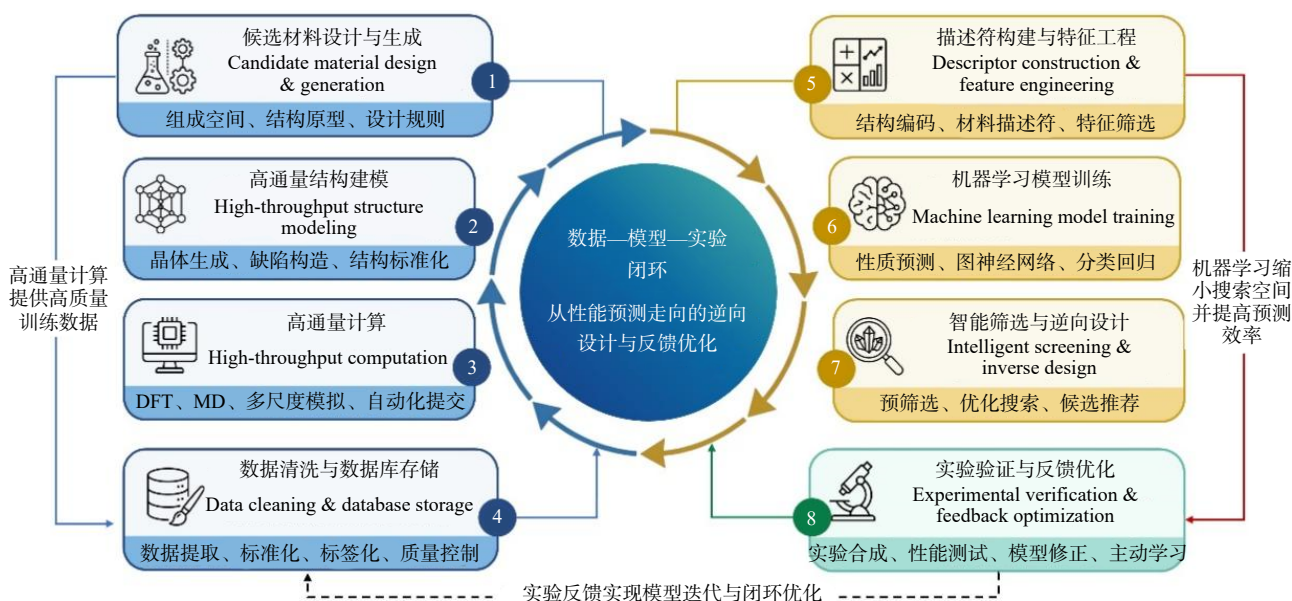


图 2 高通量计算和机器学习融合驱动材料研发的闭环框架

第一性原理计算是材料高通量计算的重要组成部分,适用于电子结构、热力学稳定性和基础物性预测;面向复杂体系和工程场景,还需结合分子动力学、相图计算(calculation of phase diagrams, CALPHAD)、相场、介观和有限元等多尺度方法,分别描述原子尺度结构演化、相稳定性、微观组织形成、宏观服役行为和工艺过程。基于密度泛函理论(density functional theory, DFT)的方法可预测电子结构、能带、态密度和弹性常数等性质,但 VASP(Vienna Ab initio Simulation Package)等程序单次计算耗时较长^[14],规模化应用依赖并行算法、高效收敛策略和自主软件开发。中国自主研发的原子轨道基第一性原理计算软件(Atomic-orbital Based Ab-initio Computation at USTC, ABACUS)同时支持平面波和数值原子轨道基组,并已与机器学习方法集成^[15-16],在保持 DFT 精度的同时

提升计算效率,为高通量计算提供支撑。

自动化 workflow 管理是高通量计算规模化应用的关键,主要承担任务调度、依赖管理、错误恢复、参数配置、数据提取和候选筛选。在多尺度场景中, workflow 还需衔接不同软件、尺度模型和数据格式,使多层次结果在统一数据框架下组织和复用,由此提高任务管理效率并降低人为操作误差。以 Materials Project 为例, pymatgen 提供从结构处理到结果分析的 Python 工具库^[17], FireWorks 支持动态 workflow 调度与容错管理^[18],二者共同构成其基础技术框架。

1.3 机器学习与高通量计算融合赋能材料研发

高通量计算虽然显著提升材料筛选效率,但仍受第一性原理计算成本制约。当设计空间和体系复杂度持续增加时,即使依托超算资源,也难以在合理时间内完成全面筛选。机器学习通过学习“成分—结构—

性能”关联,形成以代理模型替代部分 DFT 计算、以快速预筛选压缩候选空间、以大语言模型挖掘文献和数据库知识 3 条路径,分别服务模拟加速、候选筛选和知识发现。

机器学习势函数通过学习第一性原理计算得到的势能面,在特定材料体系中兼顾计算精度与模拟效率。传统分子动力学依赖经验势函数,虽能处理大尺度和长时间模拟,但精度受参数拟合质量限制,在新体系中的迁移性不足。机器学习势函数利用深度神经网络拟合构型—能量—受力关系,可在保持较高预测精度的同时提升推理效率^[19-20]。深度势能分子动力学(Deep Potential Molecular Dynamics, DeePMD)、钙钛矿氧化物通用原子间势(Universal Interatomic Potential for Perovskite Oxides, UniPero)、消息传递原子簇展开(Message Passing Atomic Cluster Expansion, MACE)等模型已在深度势能分子动力学、钙钛矿氧化物和大规模晶体体系中展示较强模拟能力^[21-25],使原本依赖大量第一性原理计算的动力学模拟、有限温度性质分析和复杂结构演化研究更具可行性。其可靠性仍受训练数据分布、构型空间覆盖程度和目标材料体系相似性约束。当模型外推至新的元素组合、缺陷结构、界面环境或反应路径时,预测误差可能显著放大。因此,这类模型更适合作为特定材料体系的高效模拟代理,而非脱离训练数据和物理验证的通用势函数,仍需主动学习、不确定性评估和第一性原理迭代校正限定适用范围。

描述符与性能预测模型主要服务大规模候选材料的快速预筛选,是高通量计算与机器学习融合的关键环节。面对庞大设计空间,即使建立高通量平台,全空间 DFT 计算仍难以实施。“先筛后算”首先利用机器学习模型对大规模候选体系快速排序和初步筛选,再将少量高潜力样本交由第一性原理计算、实验制备或高精度表征验证^[26]。其核心是构建能够表征材料组成、结构和局域化学环境的描述符,并建立描述符与目标性能之间的映射关系。描述符需要在信息完整性、计算成本和跨体系适用性之间取得平衡,既要保留与目标性质相关的结构信息,也要避免不必要的高维冗余。独立筛选与稀疏化算子(sure independence screening and sparsifying operator, SISSO)通过符号回归和压缩感知,从大规模候选特征空间筛选低

维、可解释的描述符组合并形成显式性能表达式,在热电材料、催化剂和电子结构性质预测中具有较强解释能力^[27-28]。与黑箱模型相比,SISSO 可将预测结果转化为较清晰的结构—性能关系表达,从而为后续第一性原理计算和实验验证提供更具物理含义的筛选依据。该方法的价值不仅在于提高筛选速度,还在于使模型输出能够与材料机理讨论和实验验证路径相衔接。

描述符—性能映射仍受数据来源、标签质量和模型适用域限制。计算数据通常来自理想晶体结构和标准计算条件,难以直接覆盖缺陷、界面、非平衡相和复杂工艺历史;实验数据则常受到制备方法、测试条件和服役环境差异影响。不同来源数据的误差结构和适用范围并不相同,合并建模前需要明确标签口径、测量条件和不确定性。若模型仅以组成或理想结构作为输入,容易将统计相关性误判为可迁移规律。因此,“先筛后算”主要用于压缩候选空间和提出优先验证对象,结果仍需通过第一性原理计算、实验表征和工艺验证形成闭环确认。

2 国际竞争格局与平台基础设施建设

2.1 国际材料高通量计算与机器学习发展战略

美国既是材料基因工程理念的主要提出者,也是数据驱动材料研发范式的重要引领者。2011 年启动的 MGI 明确以高通量计算、高通量实验和材料数据库协同为基础,加速新材料研发与应用,并将缩短研发周期、降低研发成本作为核心目标。此后,美国围绕数据基础设施、标准化共享、自动化 workflows 和面向制造应用的闭环创新持续完善战略部署^[29]。近年来,AI 进一步进入材料创新体系,相关布局从开放计算数据库和工具链建设,扩展到利用科研算力和高通量数据训练材料机器学习模型,推动材料发现由数据驱动筛选向 AI 辅助自主设计演进^[30]。美国由此逐步形成以开放计算数据为基础、以自动化工具链和科研算力为支撑、以 AI 模型训练和材料智能筛选为延伸的创新体系。

欧盟的战略布局更强调材料数据治理和跨平台互操作性。与美国以开放计算数据库和工具链为主要牵引的路径不同,欧盟重视可发现、可访问、可互

操作、可重用 (findable, accessible, interoperable and reusable, FAIR) 原则, 元数据规范, 材料本体, 语义技术, 以及联邦式数据基础设施。2022 年发布的《材料 2030 路线图》(《Materials 2030 Roadmap》) 将材料数字化列为优先方向, 强调以标准化语言、本体体系和可信公共数据空间提升材料数据共享与复用能力^[31]。材料设计开放数据库集成 (Open Databases Integration for Materials Design, OPTIMADE) 等机制从标准化查询、元数据规范和数据交换协议等层面推动平台互通^[32-33]。随着《科学中的人工智能战略》(《AI for Science Strategy》) 和欧洲人工智能科学资源中心 (Resource for AI Science in Europe, RAISE) 等部署推进, 欧盟正将数据治理优势延伸至 AI for Science 场景^[34], 并在材料数据长期复用、跨平台检索和跨机构协同方面形成特色及较稳定的制度安排。

英国的战略布局建立在材料科学基础研究实力和实验数据资源优势之上, 近年来逐步形成“创新战略—材料战略—AI for Science 战略”相衔接的发展框架。英国围绕国家材料战略、材料 4.0 和 AI for Science 战略, 将 AI、大数据、机器学习和数字孪生嵌入材料研发、测试与制造全链条^[35]。与美国侧重开放计算数据库、欧盟侧重数据治理规则不同, 英国更强调高质量实验数据、先进表征能力和国家算力基础设施对材料智能研发的支撑。AI 研究资源体系 (AI Research Resource, AIRR) 等建设进一步体现出英国通过主权算力和科研数据资源支撑材料基础模型与智能设计的思路。

日本在材料高通量计算与机器学习领域的战略部署, 更强调高质量实验数据、工程应用场景和产学研协同的深度结合。2015 年启动的材料信息学计划 (Materials Informatics Initiative, MII) 推动计算、实验和数据科学融合, 其特色不在于追求高通量计算数据的绝对规模, 而在于依托长期积累的高保真实验数据开展材料信息学研究^[36]。2021 年发布的《材料创新力强化战略》进一步将数据驱动确立为材料创新的重要路径, 强调整合分散数据资源、建设统一基础设施并服务产业竞争力提升^[37]。近年来, 日本将材料研发数字化纳入 Society 5.0 和统合创新战略框架, 通过相关项目强化材料数据产生、积累和活用的全生命周期管理^[38-39]。这一安排既关注数据产生环节的质量控制,

也重视数据在材料设计、工艺开发和产业协作中的持续利用。其战略重心不是单纯扩大计算数据库规模, 而是提高实验数据的可复用性, 并使材料信息学方法服务产业需求。相关部署将高质量实验数据、材料信息学方法和产业应用场景紧密连接起来, 强化日本材料研发体系中“数据产生—数据积累—数据活用—工程转化”的连续性, 并使数据驱动工具更直接地服务企业研发和工程验证。

主要经济体围绕高通量计算、材料数据库和机器学习形成不同战略组合: 美国强调开放计算数据、工具链和科研算力协同, 欧盟侧重数据治理与标准互操作, 英国依托实验数据资源和材料 4.0 体系, 日本以高质量实验数据和产业场景牵引工程转化。不同路径表明, 数据驱动材料研发的国际竞争已由单点技术能力扩展为数据质量、平台开放性、标准体系、算力供给和产业转化能力的综合竞争。

2.2 国际材料高通量计算与机器学习平台

材料数据平台与高通量计算基础设施是数据驱动材料研发的重要支撑。国际平台已形成覆盖计算数据库、实验数据库、工作流系统和标准化接口的多层次体系, 其实际价值不仅取决于数据规模, 还取决于数据来源、计算参数一致性、元数据完整性和工作流程可复现性。

美国开放计算数据库以材料项目 (Materials Project, MP)、材料发现自动化流程 (Automatic Flow for Materials Discovery, AFLOW) 和开放量子材料数据库 (Open Quantum Materials Database, OQMD) 为代表, 通过相对统一的高通量第一性原理计算流程积累晶体结构、热力学稳定性、电子结构和基础物性数据, 并与 pymatgen、matminer 等工具链共同支撑材料数据分析和机器学习建模^[40-42]。其优势在于计算流程一致、数据结构规范和开放接口较完善, 适用于大规模候选材料预筛选和机器学习模型训练, 但数据主要来自理想晶体结构和标准计算条件, 对缺陷、界面、非平衡相、制备工艺和真实服役环境覆盖有限。

欧盟平台建设更强调数据治理、科学复现和标准互通。新材料发现数据平台 (Novel Materials Discovery, NOMAD) 采用分布式联邦架构并遵循 FAIR 原则, 支持材料计算数据的全生命周期保存与统一访问^[43]; Materials Cloud 依托 AiiDA 工作流系统, 突出

计算流程可追溯与结果复现^[44]; OPTIMADE 通过统一查询接口增强不同数据库的可检索和可复用能力^[45]。这类平台更重视数据长期复用、跨平台访问和科学复现,其可靠运行仍取决于用户数据完整性、元数据标注规范,以及异构数据整合与质量控制。

英国和日本的平台体系更突出实验数据积累和工程应用导向。剑桥结构数据库(Cambridge Structural Database, CSD)长期收录经严格审核的晶体结构数据,在晶体工程、药物分子设计和功能材料研究中提供支撑;日本 NIMS 材料数据库(NIMS Materials Database, MatNavi)积累多类材料实验测量数据,Starry-data2 等工具进一步支持文献图表数据提取、实验数据组织和材料信息学分析^[32,46]。相较于计算数据库,实验数据库更接近真实材料体系和应用场景,但也更易受测试条件、样品制备历史、数据格式和元数据完整性差异影响。用于跨体系机器学习建模时,仍需数据清洗、标准化处理和不确定性评估。

开放计算数据库、数据治理与 workflow 平台、实验数据库分别强化规模化预测、流程复现和高保真实验校准能力。材料研发走向 AI 辅助设计和闭环优化后,平台建设重点已由数据规模扩张转向数据能否被可靠调用、准确追溯和有效验证。

2.3 中国材料高通量计算与机器学习发展战略

高通量计算、材料数据库与机器学习的融合已纳入中国新材料创新体系建设。围绕缩短研发周期、降低试错成本和提升核心工具自主可控能力,国家持续以材料基因工程、数据基础设施和 AI 赋能推进制度部署。“十三五”材料基因工程重点专项由此奠定了中国数据驱动研发的基础框架。早期国内研究已从高通量计算、材料数据库和数据挖掘等方面讨论计算与数据方法的基础作用,提出通过计算、数据和实验协同改变传统试错式研发模式^[47]。2015 年设立的“材料基因工程关键技术与支撑平台”重点专项围绕缩短研发周期、降低研发成本,部署高通量计算、高通量实验和数据库 3 类支撑平台;2016 年相关规划进一步将材料基因工程列为重点方向。这一阶段重点构建材料数据生成、管理和应用的基础能力,使国内研发在平台、流程和数据资源层面具备协同基础,并推动分散式计算和实验探索逐步转向高通量方法与数据库共同支撑的组织方式。

进入“十四五”时期,相关部署由研发前端扩展到工程验证和产业转化。《“十四五”原材料工业发展规划》提出推进材料基因工程、建设材料数字化研发平台,并推动其向新材料研发、生产和应用全过程延伸;《新材料中试平台建设指南(2024—2027 年)》进一步强调中试平台与材料大数据中心、材料基因工程平台协同。由此,材料基因工程不再仅对应候选筛选和性能预测,而是进一步延伸至中试验证、工艺放大与应用评价。

国家级数据基础设施建设构成战略演进的重要支点。2024 年印发的新材料大数据中心总体建设方案,将其定位为促进新材料产业创新发展的新型研发基础设施,并提出“1+N”架构^[48]。这一部署把材料数据资源组织、算法工具供给和应用示范纳入国家级基础设施框架,使材料创新战略由单点平台建设进一步转向体系化协同。2025 年,国务院“人工智能+”行动提出实施“人工智能+科学技术”和“人工智能+产业发展”^[49];北京“人工智能+新材料”行动计划则提出开发新材料智能研发垂类模型与软件系统^[50]。2 项部署分别从国家行动和区域实践层面明确 AI 进入材料研发与产业应用的方向。相关政策进一步推动 AI 由通用赋能工具进入材料设计、工艺优化和制造应用场景,并促进数据资源、模型工具与应用示范衔接,使材料数据基础设施、垂类模型和产业任务逐步形成相互支撑的体系。

中国已形成材料基因工程、材料数据基础设施和 AI 应用共同支撑的发展框架,后续仍需完善高质量实验数据、数据标准与共享机制、自主核心软件和自动化 workflow 生态,加强“计算—实验—中试—制造”链条衔接,并将分散的数据资源、计算工具、模型算法和产业场景连接为可复用、可验证、可迭代的研发闭环,以提升材料研发由第四范式向第五范式演进的实际成效。

2.4 中国材料高通量计算与机器学习平台建设进展

在材料基因工程专项和国家级数据基础设施建设推动下,中国材料高通量计算与机器学习平台建设已转向体系化布局,初步形成涵盖自主计算软件、workflow 平台、机器学习势函数、材料数据库、智能设计平台和材料大模型的技术体系(图 3、表 1)。

自主计算软件与高通量计算和 workflow 平台构成

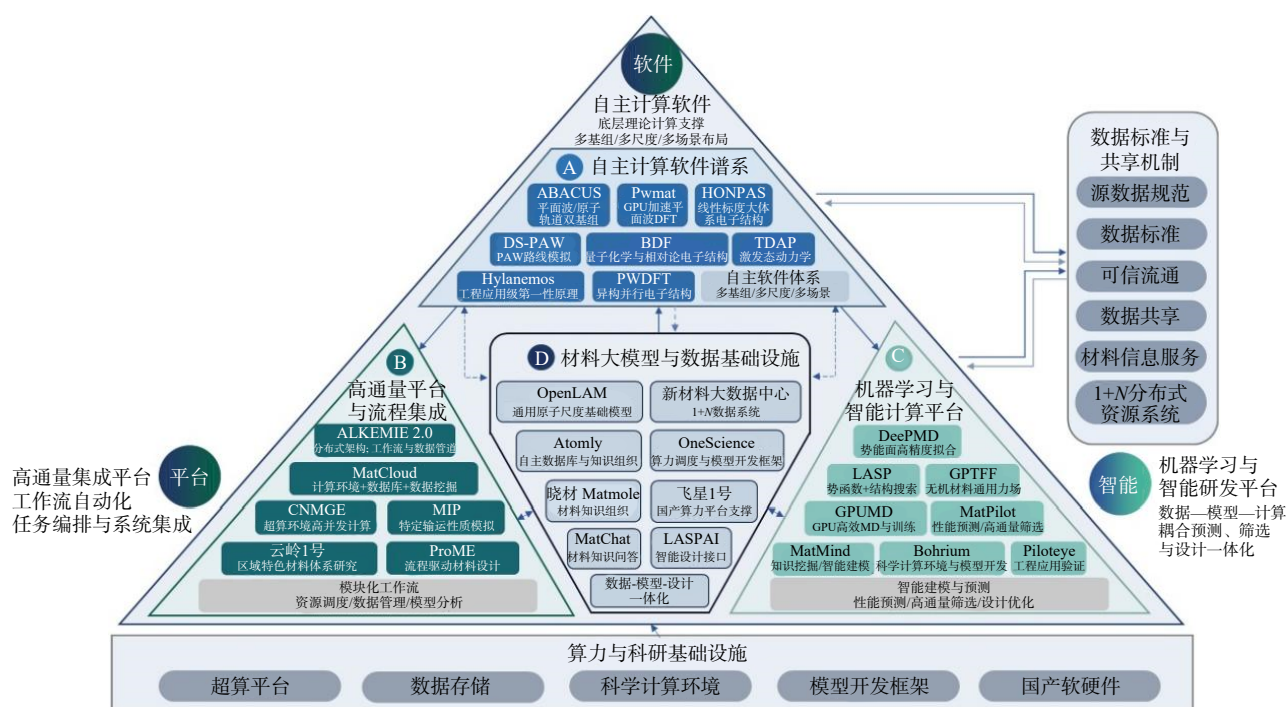


图3 中国材料高通量计算与机器学习平台生态体系

材料数据生成的前端基础。ABACUS、PWmat、DS-PAW等自主或国产化电子结构软件,以及ALKEMIE、MatCloud等高通量计算与材料信息学平台,已覆盖相关计算和工作流环节^[16,51-52]。平台应用成效取决于稳定、可复现和可扩展的数据生成流程;跨软件接口、统一数据格式、自动化纠错、工作流复现和长期生态维护,是进一步提升持续应用能力的重要方向。

机器学习势函数和智能模拟工具进一步拓展了材料模拟的时空尺度。DeePMD、OpenLAM和GPTFF等工具分别从深度势能模型训练、通用原子尺度基础模型和大模型思想融入势函数构建等方向提升复杂体系模拟能力^[22,53-57]。其可靠性仍受训练数据覆盖范围、材料体系相似性和分布外不确定性评估制约。面对新的元素组合、缺陷结构或界面环境,相关模型仍需依托主动学习、第一性原理计算迭代校正和实验结果验证限定适用范围,不能简单等同于具有普适迁移能力的通用模拟工具。

材料数据库、知识平台、材料大模型和AI设计平台正推动国内体系由数据存储向知识与决策支撑延伸。新材料大数据中心和Atomly承担材料数据汇聚、结构—性能检索、文献数据组织与知识服务^[48,58];MatPilot、MatMind、MatChat和LASPAI探索材料知识问答、候选材料生成、结构—性能预测、实验方案

设计和科研流程辅助^[59-62]。上述平台能否支撑可靠发现,取决于数据来源可追溯、元数据完整、计算流程可复现、模型输出可追溯,以及候选结果能否进入实验验证和工艺优化闭环。高质量实验数据、负结果记录和应用场景反馈的持续积累,可推动平台由知识问答和候选推荐进一步延伸至可验证的材料设计。Bohrium作为科学计算环境和算力基础设施代表,为高通量计算、模型训练、任务管理和可复现研究环境提供底层支撑^[63]。随着国内平台类型增加,建设重点应由单个平台功能拓展转向跨平台数据标准、接口协同、开放服务机制和长期维护能力,推动“计算—数据—模型—实验—中试—制造”连续闭环贯通并形成材料智能研发基础设施体系。

3 数据驱动的新材料产业变革

当前,全球新材料产业正经历从“经验试错”到“数据预测”的变革,材料研发的技术路径、产业生态和竞争格局随之改变;在中国发展新质生产力的背景下,数据驱动研发正由科研探索走向产业应用,成为服务关键材料自主可控和产业升级的重要路径,以下据此讨论其对新质生产力、产业生态及重点材料应用的影响,并分析数据质量、工艺放大和应用验证对产

表 1 中国代表性材料高通量计算与机器学习软件/平台

平台名称	开发单位	平台类型	主要功能
ABACUS	中国科学技术大学、北京大学、中国科学院物理研究所、北京科学智能研究院等	第一性原理计算软件	支撑电子结构计算、结构优化、分子动力学模拟和高通量第一性原理数据生成
PWmat	北京龙讯旷腾科技有限公司	第一性原理计算软件	支撑图形处理器(GPU)加速平面波DFT计算、能带结构、缺陷、界面和输运性质模拟
DS-PAW	鸿之微科技	第一性原理计算软件与材料模拟平台	基于投影缀加波(PAW)方法支撑块体、表面、界面、低维材料和液体体系模拟
ALKEMIE	北京航空航天大学孙志梅团队	高通量计算与材料信息学平台	集成高通量自动流程、数据管理、机器学习分析和候选材料筛选
MatCloud	北京迈高材云科技有限公司	高通量计算与数据管理平台	高通量计算驱动引擎;图形化界面;材料计算数据库;云端自动化流程
DeePMD-kit	DeepModeling/Deep Potential开源社区	机器学习势函数与分子动力学	支撑深度势能模型训练、分子动力学模拟和复杂体系原子尺度模拟
GPUMD	GPUMD开源社区等	GPU分子动力学与机器学习势函数工具	支撑GPU加速分子动力学、机器学习势函数训练和大规模原子模拟
OpenLAM	北京科学智能研究院、DeepModeling社区等	大原子模型平台	面向材料体系构建通用原子尺度基础模型,支撑机器学习势函数和原子尺度模拟
Atomly	中国科学院物理研究所、松山湖材料实验室等	材料数据库与知识平台	支撑无机材料数据管理、DFT计算数据组织、结构—性能检索和材料知识服务
晓材 Matmole	晓材Matmole团队	材料数据库与知识服务平台	支撑材料数据检索、金属牌号、相图、结构—性能数据和文献知识组织
新材料大数据中心	工业和信息化部、财政部、国家数据局等	国家级材料数据基础设施	支撑材料数据汇聚、共享服务、资源节点建设、算法工具和应用示范
MatChat	松山湖材料实验室、中国科学院计算机网络信息中心等	材料大语言模型与应用平台	支撑材料知识问答、无机材料合成路径推理和材料知识生成
MatPilot	国防科技大学	材料大模型与多智能体平台	支撑人机协同、多智能体协作、候选材料生成和实验方案设计
MatMind	中国科学院上海硅酸盐研究所	材料科学大模型与AI平台	支撑材料知识推理、结构—性能预测、多专家协同和智能实验平台建设
Bohrium	深势科技	科学计算与AI for Science平台	支撑科学计算、任务管理、模型开发、算力调度和可复现研究环境

业转化的约束。

3.1 新质生产力发展

新材料是战略性新兴产业的重要基础,其发展水平关系国家科技竞争力与产业安全。高通量计算与机器学习融合形成的数据驱动研发体系,使材料设计由经验试错转向模型预测与人机协同,材料体系由实体样品扩展为数据化表征,科研主体也更多承担模型设计与系统调控任务。计算、实验和应用持续生成数据,数据再反向提升模型能力,形成“数据—模型—应用”反馈闭环。该闭环的产业效能取决于预测能否通过实验结果、工艺参数和真实服役反馈持续校正;在相关数据和反馈不足时,模型主要用于候选推荐与性能预估,仍需实验和工艺验证才能形成可制造、可规模化的材料方案。

3.2 产业生态重构:从“经验驱动”到“数据驱动”

传统材料研发依赖经验积累和单点突破,关键知识附着于个体经验,实验数据分散于机构和项目,难以形成可复用资产。数据驱动范式使材料研发由分散试错转向系统协同。产业数据除成分、结构和性能外,还需覆盖制备方法、工艺参数、测试条件、批次差异、失效模式和服役环境等链式信息。只有这些信息被系统记录和标准化组织,机器学习模型才能由局部相关性分析进一步进入工艺反馈和质量控制环节。材料数据基础设施和流通机制可推动分散数据转化

为可共享、可治理的资源,国家新材料大数据中心等平台通过分布式体系整合跨区域、跨机构数据,相关数据要素政策和安全流通技术也为开放共享与跨主体协同提供支撑。产业应用还要求数据来源可追溯、元数据完整、测试条件明确和格式可复用,材料数据能否支撑可靠建模、工艺反馈和质量评价,将直接影响数据驱动方法的实际作用。

3.3 重点新材料产业变革

高通量计算与机器学习正在改变重点新材料的

研发组织方式,其主要作用不是替代实验,而是通过候选生成、模型预测、排序和参数优化压缩设计空间,将有限的计算与实验资源集中到更具验证价值的体系。其有效性既取决于材料数据库、计算流程和模型能否提供可靠判断,也取决于结果能否进入实验验证、工艺放大、器件测试和服役评价。新能源、催化、结构及先进无机非金属、电子与半导体材料已展示相应应用路径(图4)。

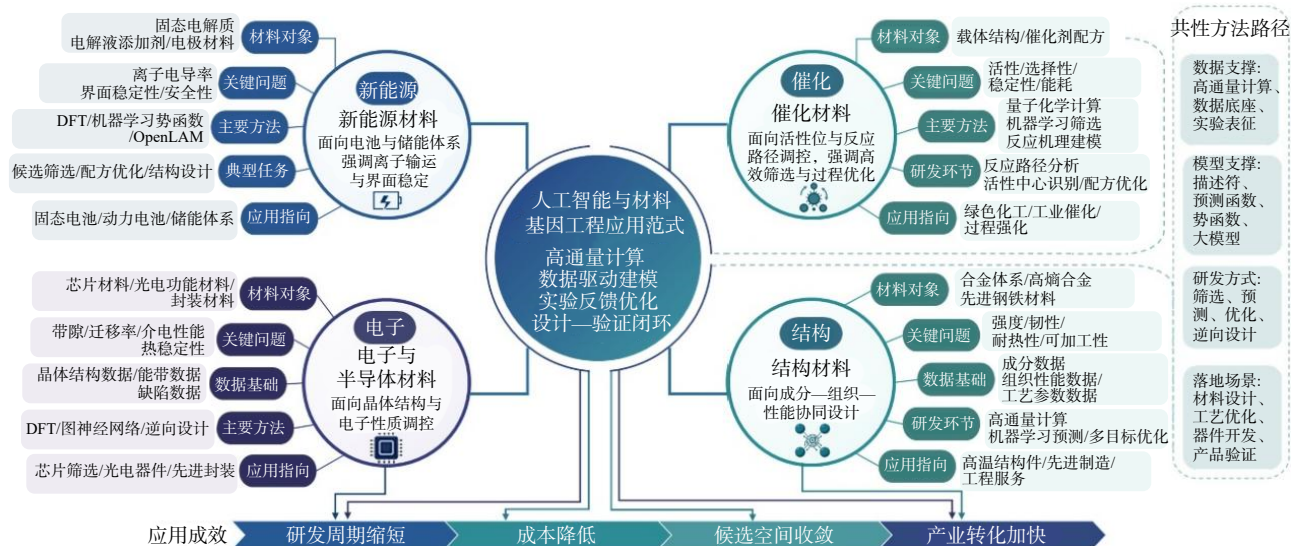


图4 数据驱动材料研发在重点新材料领域的应用路径

新能源材料中,电池体系通常涉及溶剂、锂盐、功能添加剂、电极界面和器件结构等多维变量,产业研发目标也不只限于容量或能量密度提升,还包括循环寿命、安全性、界面稳定性和制造一致性。面向性能、安全与效率协同优化,传统经验筛选和逐步实验验证需与数据驱动方法互补。相关方法可利用历史实验、计算结果和器件测试数据识别变量关联,据此确定优先验证的组成和工艺窗口。在无机固态电解质研究中,机器学习辅助的分层筛选可从大规模候选体系中优先识别潜在材料,并通过后续计算或实验验证缩小搜索空间^[64]。该方法提高了初筛效率,但候选体系进入实际应用仍取决于电极—电解质界面稳定性、电芯加工一致性、长循环安全性和规模化制造重复性验证。

催化材料研发需协同优化活性、选择性、稳定性和成本,其性能受活性中心、载体结构、反应路径和

操作条件共同影响,传统实验筛选周期长且变量耦合强。高通量计算与机器学习可将材料结构、局域化学环境、反应特征和选择性信息纳入统一筛选框架,提高候选催化剂识别效率。以 CO_2 环加成反应金属有机框架(metal-organic framework, MOF)催化剂为例,高通量筛选结合可解释机器学习模型,可分析候选MOF的结构特征与催化性能关系,提高发现效率和结果解释性^[65]。相关分析还可识别影响性能的关键因素,为后续实验设计提供依据。面向产业应用,模型结果仍需在真实反应条件下检验,实验放大、长周期稳定性和装置运行验证仍是工程转化的关键。

结构及先进无机非金属材料更强调多性能协同和服役可靠性,强度、韧性、耐热性、介电与压电性能、加工窗口和成本往往相互制约,相关性能还受到组成、微观组织、缺陷分布、晶界结构和工艺历史显著影响。机器学习可在成分、组织、工艺和性能之间

建立映射,并通过可解释性分析识别关键结构因素。对于此类材料,模型还需辅助寻找多目标约束下的候选区域。压电材料研究中,图机器学习与知识图谱已用于高通量筛选新型候选体系,为无机功能材料的组成设计和性能预测提供数据驱动路径^[66]。该方法可提高候选筛选和结构—性能关系解析效率,但工程应用仍受制备、加工、相稳定性和批次一致性制约,候选材料需进一步经过加工放大、环境稳定性、器件或构件性能及长期可靠性验证。

电子与半导体材料对带隙、载流子迁移率、介电常数、缺陷形成能、界面态和器件稳定性等指标要求严格,材料筛选需统筹电子结构、制备工艺和器件应用条件。高通量计算与机器学习可从大规模候选中快速识别满足特定功能指标的体系,并为薄膜制备、界面调控和器件设计提供优先验证对象。在二维电子材料中,相关方法已用于范德华介电材料筛选和分类,以识别适用于二维纳米电子器件的候选介质^[67]。这使电子材料设计由少量已知体系扩展到更大候选空间。候选材料进入器件应用仍需解决薄膜生长窗口、掺杂可控性、接触电阻、缺陷态、界面稳定性和长期可靠性问题。

总体而言,数据驱动方法的功能是前移候选生成、性能预测和验证优先级排序,而非直接给出最终材料配方。其产业价值取决于模型输出能否与高质量实验数据、可复现计算流程、明确工艺参数和真实服役反馈形成闭环。高通量计算与机器学习能够加快材料发现和优化,但候选方案进入产业体系仍需经过实验验证、工艺放大、器件或构件评价、长期稳定性和成本约束分析。数据驱动研发因而更适合作为决策加速和研发资源配置工具,其作用仍需与物理机制、实验验证和制造条件结合。

4 结论与展望

高通量计算与机器学习的融合,正在推动材料研发由经验试错、理论分析和单点计算转向数据、模型与实验协同的新范式。其意义不仅是提升筛选效率,更在于压缩设计空间、前移性能预测环节并形成“计算—预测—验证—反馈”闭环,推动材料科学由数据密集型研究迈向智能化研究。全球主要经济体已围

绕材料数据基础设施、标准体系、科研算力和AI平台展开布局,材料创新竞争由单一技术能力扩展为数据、工具、平台和组织体系的综合竞争。中国自“十三五”以来持续推进材料基因工程、材料大数据中心和“人工智能+制造”,已形成计算软件、高通量平台、数据资源和材料大模型等体系,但核心软件自主化、高质量数据、标准体系及研发—产业贯通仍需加强。

面向下一阶段,材料智能研发的发展重点可归纳为3个方面:一是继续完善国家级材料数据基础设施,提升数据汇聚、治理、共享和复用能力,形成稳定、可持续的数据供给体系;二是加快国产计算软件、高通量平台和材料大模型协同发展,增强关键工具链的自主可控能力;三是强化研发、中试和产业化衔接,推动数据驱动方法由候选筛选和性能预测进一步延伸至工艺优化、器件设计和规模化制造,提升关键材料的工程转化效率和产业应用能力。

参考文献(References)

- [1] 祝君璧. 新材料产业迈向质的飞跃[N]. 经济日报, 2022-06-17(9).
- [2] Tolle K M, Tansley D S W, Hey A J G. The fourth paradigm: Data-intensive scientific discovery[J]. *Proceedings of the IEEE*, 2011, 99(8): 1334-1337.
- [3] Ramprasad R, Batra R, Piliand G, et al. Machine learning in materials informatics: Recent applications and prospects[J]. *npj Computational Materials*, 2017, 3: 54.
- [4] 关永军, 陈柳, 王金三. 材料基因组技术内涵与发展趋势[J]. *航空材料学报*, 2016, 36(3): 71-78.
- [5] 温李阳, 姚明佳, 陈忻. 智能驱动的材料研发态势及对策建议[J]. *前瞻科技*, 2025, 4(1): 16-27.
- [6] 代梦艳, 刘喆, 彭文联, 等. 人工智能加速材料发现与生成[J]. *防化研究*, 2022, 1(1): 53-58.
- [7] 李国杰. 智能化科研(AI4R): 第五科研范式[J]. *中国科学院院刊*, 2024, 39(1): 1-9.
- [8] 鄂维南. AI助力打造科学研究新范式[J]. *中国科学院院刊*, 2024, 39(1): 10-16.
- [9] 张平文. AI4S时代, 应用数学发展的新机遇[EB/OL]. [2026-06-29]. <https://m.csiam.org.cn/1035/202209/499.html>.
- [10] 余江, 张越, 周易. 人工智能驱动的科研新范式及学科应用研究[J]. *中国科学院院刊*, 2025, 40(2): 362-370.
- [11] Zhang X, Wang L M, Helwig J, et al. Artificial intelligence for science in quantum, atomistic, and continuum systems[J]. *Foundations and Trends in Machine Learning*, 2025, 18(4): 385-849.

- [12] Merchant A, Batzner S, Schoenholz S S, et al. Scaling deep learning for materials discovery[J]. *Nature*, 2023, 624(7990): 80–85.
- [13] 杨小渝. 高通量多尺度材料计算和机器学习[M]. 北京: 科学出版社, 2023.
- [14] Kresse G, Furthmüller J. Efficient iterative schemes for *ab initio* total-energy calculations using a plane-wave basis set[J]. *Physical Review B*, 1996, 54(16): 11169–11186.
- [15] Lin P Z, Ren X G, He L X. Strategy for constructing compact numerical atomic orbital basis sets by incorporating the gradients of reference wavefunctions[J]. *Physical Review B*, 2021, 103(23): 235131.
- [16] 陈默涵. 密度泛函理论软件ABACUS进展及其与深度学习算法的融合及应用[J]. *金属学报*, 2024, 60(10): 1405–1417.
- [17] Ong S P, Richards W D, Jain A, et al. Python Materials Genomics (pymatgen): A robust, open-source Python library for materials analysis[J]. *Computational Materials Science*, 2013, 68: 314–319.
- [18] Jain A, Ong S P, Chen W, et al. FireWorks: A dynamic workflow system designed for high-throughput applications[J]. *Concurrency and Computation: Practice and Experience*, 2015, 27(17): 5037–5059.
- [19] Behler J, Parrinello M. Generalized neural-network representation of high-dimensional potential-energy surfaces[J]. *Physical Review Letters*, 2007, 98(14): 146401.
- [20] Bartók A P, Payne M C, Kondor R, et al. Gaussian approximation potentials: The accuracy of quantum mechanics, without the electrons[J]. *Physical Review Letters*, 2010, 104(13): 136403.
- [21] Zhang L F, Han J Q, Wang H, et al. Deep potential molecular dynamics: A scalable model with the accuracy of quantum mechanics[J]. *Physical Review Letters*, 2018, 120(14): 143001.
- [22] Wang H, Zhang L F, Han J Q, et al. DeePMD-kit: A deep learning package for many-body potential energy representation and molecular dynamics[J]. *Computer Physics Communications*, 2018, 228: 178–184.
- [23] Wu J, Yang J Y, Liu Y J, et al. Universal interatomic potential for perovskite oxides[J]. *Physical Review B*, 2023, 108(18): L180104.
- [24] Batatia I, Kovacs D P, Simm G, et al. MACE: Higher order equivariant message passing neural networks for fast and accurate force fields[C]//Proceedings of Advances in Neural Information Processing Systems 35. Neural Information Processing Systems Foundation, Inc. (NeurIPS), 2022: 11423–11436.
- [25] Batatia I, Benner P, Chiang Y, et al. A foundation model for atomistic materials chemistry[J]. *The Journal of Chemical Physics*, 2025, 163(18): 184110.
- [26] Schmidt J, Marques M R G, Botti S, et al. Recent advances and applications of machine learning in solid-state materials science[J]. *npj Computational Materials*, 2019, 5: 83.
- [27] Schütt K T, Glawe H, Brockherde F, et al. How to represent crystal structures for machine learning: Towards fast prediction of electronic properties[J]. *Physical Review B*, 2014, 89(20): 205118.
- [28] Ouyang R H, Curtarolo S, Ahmetcik E, et al. SISSO: A compressed-sensing method for identifying the best low-dimensional descriptor in an immensity of offered candidates[J]. *Physical Review Materials*, 2018, 2(8): 083802.
- [29] 张帆, 杨艳, 刘雨虹. “人工智能+新材料”领域发展现状、趋势和启示[J]. *国际石油经济*, 2025, 33(9): 19–25, 85.
- [30] 朱微微. 美国构建AI4S科研生态的路径研究及启示[J]. *世界科技研究与发展*, 2026, 48(2): 168–179.
- [31] European Union. Materials 2030 roadmap[EB/OL]. [2026–03–29]. <https://www.ami2030.eu/roadmap/>.
- [32] 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): 89–102.
- [33] 张澜庭, 刘明洋, 袁宏建, 等. AI驱动的新材料智能研发与数据标准化[J]. *科学通报*, 2025, 70(24): 4066–4080.
- [34] 于佳佳, 晁嫚璐. 欧美AI for Science战略布局动向与启示[J]. *科技中国*, 2026(1): 24–31.
- [35] 王硕, 高睿婧, 李正凤. 全球AI for Science战略动向及对我国的启示[J]. *科学学研究*, 2026, 44(3): 449–463, 528.
- [36] National Institute for Materials Science. Materials research by information integration initiative[EB/OL]. [2026–07–12]. <https://www.nims.go.jp/MII-I/en/>.
- [37] Cabinet Office, Government of Japan. Materials innovation strategy[EB/OL]. (2021–04–27) [2026–07–12]. https://www8.cao.go.jp/cstp/material/material_honbun_en.pdf.
- [38] Ministry of Education, Culture, Sports, Science and Technology, Japan. Materials DX Platform Initiative[EB/OL]. (2024–03–07) [2026–07–12]. https://www.mext.go.jp/content/20240307-mxt-mxt_nanozai-000034134_3.pdf.
- [39] Cabinet Office, Government of Japan. Integrated Innovation Strategy 2025[EB/OL]. [2026–07–12]. https://www8.cao.go.jp/cstp/tougosenryaku/togo2025_honbun_eiyaku.pdf.
- [40] Curtarolo S, Setyawan W, Hart G L W, et al. AFLOW: An automatic framework for high-throughput materials discovery[J]. *Computational Materials Science*, 2012, 58: 218–226.
- [41] 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.
- [42] Saal J E, Kirklin S, Aykol M, et al. Materials design and

- discovery with high-throughput density functional theory: The open quantum materials database (OQMD)[J]. *JOM*, 2013, 65(11): 1501–1509.
- [43] Scheidgen M, Himanen L, Ladines A N, et al. NOMAD: A distributed web-based platform for managing materials science research data[J]. *Journal of Open Source Software*, 2023, 8(90): 5388.
- [44] Pizzi G, Cepellotti A, Sabatini R, et al. AiiDA: Automated interactive infrastructure and database for computational science[J]. *Computational Materials Science*, 2016, 111: 218–230.
- [45] 封俊杰, 周琨, 李木琛, 等. 数据驱动研究范式下材料数据库的构建与应用[J]. *科学通报*, 2025, 70(24): 4044–4065.
- [46] Katsura Y, Kumagai M, Mato T, et al. Starrydata: From published plots to shared materials data[J]. *Science and Technology of Advanced Materials: Methods*, 2025, 5(1): 2506976.
- [47] 杨小渝, 任杰, 王娟, 等. 基于材料基因组计划的计算和数据方法[J]. *科技导报*, 2016, 34(24): 62–67.
- [48] 工业和信息化部, 财政部, 国家数据局. 关于印发《新材料大数据中心总体建设方案》的通知(工信部联原〔2024〕205号)[EB/OL]. (2024–10–16) [2026–07–23]. <https://app.www.gov.cn/govdata/gov/202410/31/521191/article.html>.
- [49] 国务院. 国务院关于深入实施“人工智能+”行动的意见(国发〔2025〕11号)[EB/OL]. (2025–08–21) [2026–07–23]. https://www.gov.cn/zhengce/content/202508/content_7037861.htm.
- [50] 北京市科学技术委员会、中关村科技园区管理委员会, 北京市发展和改革委员会, 北京市经济和信息化局. 北京市加快推动“人工智能+新材料”创新发展行动计划(2025—2027年)[EB/OL]. (2025–01–26) [2026–07–23]. https://www.beijing.gov.cn/zhengce/zhengcefaui/202501/t20250127_4000881.html.
- [51] Wang G J, Peng L Y, Li K Q, et al. ALKEMIE: An intelligent computational platform for accelerating materials discovery and design[J]. *Computational Materials Science*, 2021, 186: 110064.
- [52] Yang X Y, Wang Z G, Zhao X S, et al. MatCloud, a high-throughput computational materials infrastructure: Present, future visions, and challenges[J]. *Chinese Physics B*, 2018, 27(11): 110301.
- [53] Zeng J Z, Zhang D, Lu D H, et al. DeePMD-kit v2: A software package for deep potential models[J]. *The Journal of Chemical Physics*, 2023, 159(5): 054801.
- [54] Fan Z Y, Wang Y Z, Ying P H, et al. GPUMD: A package for constructing accurate machine-learned potentials and performing highly efficient atomistic simulations[J]. *The Journal of Chemical Physics*, 2022, 157(11): 114801.
- [55] Zhang D, Liu X, Zhang X Y, et al. DPA-2: A large atomic model as a multi-task learner[J]. *npj Computational Materials*, 2024, 10: 293.
- [56] Xie F K, Lu T L, Meng S, et al. GPTFF: A high-accuracy out-of-the-box universal AI force field for arbitrary inorganic materials[J]. *Science Bulletin*, 2024, 69(22): 3525–3532.
- [57] Peng A Y, Liu X, Guo M Y, et al. The OpenLAM challenges: LAM crystal philately competition[J]. *Machine Learning: Science and Technology*, 2025, 6(2): 020701.
- [58] Liu M, Meng S. Atomly. net materials database and its application in inorganic chemistry[J]. *Scientia Sinica Chimica*, 2023, 53(1): 19–25.
- [59] Ni Z Q, Li Y H, Hu K J, et al. MatPilot: An LLM-enabled AI materials scientist under the framework of human-machine collaboration[EB/OL]. (2024–11–10) [2026–07–12]. <https://arxiv.org/abs/2411.08063>.
- [60] 中国科学院上海硅酸盐研究所. 上海硅酸盐所发布材料科学与工程大模型MatMind [EB/OL]. (2025–04–07) [2026–07–12]. https://www.sic.cas.cn/xwzx/tpxw/202504/t20250407_7589586.html.
- [61] Chen Z Y, Xie F K, Wan M, et al. MatChat: A large language model and application service platform for materials science[J]. *Chinese Physics B*, 2023, 32(11): 118104.
- [62] 罗涵之, 梁琦茗, 郭子兴, 等. LASPAI: 人工智能驱动的未来原子模拟平台[J]. *物理化学学报*, 2026, 42(6): 100235.
- [63] 深势科技. Bohrium微尺度科学计算云平台[EB/OL]. [2026–03–29]. <https://bohrium.dp.tech>.
- [64] Chen W J, Zhou J H, Li S T, et al. Accelerated discovery of novel inorganic solid-state electrolytes through machine learning-assisted hierarchical screening[J]. *Journal of Alloys and Compounds*, 2025, 1010: 177981.
- [65] Bai X F, Li Y, Xie Y B, et al. High-throughput screening of CO₂ cycloaddition MOF catalyst with an explainable machine learning model[J]. *Green Energy & Environment*, 2025, 10(1): 132–138.
- [66] Anand A, Kumari P, Kalyani A K. High throughput screening of new piezoelectric materials using graph machine learning and knowledge graph approach[J]. *Computational Materials Science*, 2025, 246: 113445.
- [67] Li Y H, Wan G L, Zhu Y Q, et al. High-throughput screening and machine learning classification of van der Waals dielectrics for 2D nanoelectronics[J]. *Nature Communications*, 2024, 15: 9527.

A new data-driven paradigm for materials research and design: Integrating high-throughput computing and machine learning

SA Baisheng¹, CHEN Lintao¹, ZHU Linggang², YANG Mingli³, SU Yanjing⁴, ZHOU Jian², SUN Zhimei^{2*}, XIE Jianxin⁴

1. College of Materials Science and Engineering, Fuzhou University, Fuzhou 350108, China

2. School of Materials Science and Engineering, Beihang University, Beijing 100083, China

3. Research Center for Materials Genome Engineering, Sichuan University, Chengdu 610065, China

4. School of Advanced Materials Innovation, University of Science and Technology Beijing, Beijing 100083, China

Abstract The rapid development of artificial intelligence is driving a new wave of scientific and technological revolution. In materials science, it is promoting the transformation of data-driven materials R&D systems, represented by Materials Genome Engineering, from an experience-driven model toward an integrated mode combining data-driven approaches with intelligent design. Focusing on the paradigm shift in materials R&D driven by the integration of high-throughput computing and machine learning, this review systematically summarizes the evolution of materials R&D paradigms, the key technological systems of high-throughput computing and machine learning, strategic layouts and progress in platform construction in China and abroad, and analyzes their impacts on the ecosystem of the advanced materials industry, scientific research organization models, and R&D pathways for key materials. High-throughput computing supports materials data generation and large-scale screening, while machine learning facilitates closed-loop iteration in intelligent materials design through rapid pre-screening and knowledge discovery, shifting competition in materials innovation from individual technological capabilities to the integrated capabilities of data resources, computational tools, R&D platforms, and organizational systems. China has initially established a relevant foundation and competitiveness in this field, and future efforts may further focus on improving the autonomy of core software, accumulating high-quality data, refining standard systems, and strengthening the integration of the R&D chain with the industrial chain. The deep integration of high-throughput computing and machine learning is driving the evolution of materials R&D from the fourth paradigm to the fifth paradigm and is continuously reshaping the technological system, industrial ecosystem, and research organization logic of advanced materials innovation.

Keywords materials genome engineering; high-throughput computation; machine learning; materials R&D paradigm; materials database ●



(责任编辑 魏来)