

AI 驱动的新材料智能研发与数据标准化

张澜庭^{1,2}, 刘明洋¹, 袁宏建^{1,2,3}, 汪洪^{1,2,3,4*}

1. 上海交通大学材料科学与工程学院, 上海 200240

2. 上海交通大学材料基因组联合研究中心, 上海 200240

3. 上海交通大学张江高等研究院, 上海 201203

4. 苏州实验室, 苏州 251123

* 联系人, E-mail: hongwang2@sjtu.edu.cn

2025-03-31 收稿, 2025-06-20 修回, 2025-06-30 接受, 2025-07-01 网络版发表

摘要 近年来, 随着人工智能 (artificial intelligence, AI) 和自动化技术的快速发展, AI 驱动的自主实验室在新材料智能研发领域展现出巨大潜力, 将成为新材料研发范式变革的“新基建”。本文聚焦 AI 驱动的自主实验室在加速新材料发现中的国内外现状和核心挑战, 综述了 AI 驱动自主实验系统在加速新材料发现中的最新进展。自主实验室通过将实验室自动化、机器人技术和 AI 算法、数据库融合为一个整体, 形成闭环反馈 workflow, 在无需人工干预下高效优化目标性能。根据自主实验室硬件和软件的技术特征和自主化程度, 可将其分为等级 0 至等级 5。数据驱动是 AI 技术的基础, 除了自主化程度外, 自主实验室具备智能化数据工厂的特征, 自主实验室既是数据的生产者也是数据的使用者, 数据标准化是打破自主实验室信息孤岛, 从“孤立智能”迈向“协同智能”的关键环节。文中最后探讨了自主实验室建设面临的挑战并对未来发展方向进行了展望。

关键词: 人工智能、自主实验、材料设计、数据标准、数据驱动

传统的“试错模式”已经难以满足新材料研发的需求。近年来, 人工智能 (artificial intelligence, AI) 技术和自动化高通量实验的最新进展正推动新材料研发进入智能化模式。自动化实验流程起源于 20 世纪 40 年代, Bosch 和 Mittasch 等人^[1~3]设计出了高度可并行化的机器人连续流平台, 对 Haber-Bosch 工艺的固氮催化剂进行网格式搜索。通过 20000 次实验, 测试了 4000 种化合物, 筛选出一种基于铁氧化物的混合催化剂, 不仅性能接近昂贵且稀缺的钨催化剂, 而且成本更低、更易于大规模生产。1990 年代出现的高通量实验技术 (如组合材料芯片技术^[4~7]等) 通过同位并行的方法^[8]提高了材料研发效率。然而, 在材料研发的高维参数优化中仍旧存在大量人力和物力资源的浪费, 需要在实验过程中通过对实验计划的实时优化来解决。Isenhour^[9]于 1985 年首次讨论了在实验中使用人工智能规划机器人的问题, 对智能化实

验进行了大胆前瞻，但限于当时条件，未能获得即时的效果。进入 21 世纪，随着 AI^[10~13]与机器学习（machine learning, ML）^[14~16]方法被引入实验流程，使得实验系统能够实现自动化、智能化甚至自主化^[15,17~20]，对自驱动实验系统（self-driving laboratory, SDL）的出现起到了至关重要的推动作用。自主实验正在成为未来科学研究的重要手段^[21]。SDL 将实验室自动化（例如合成^[22]、分离^[23]、纯化^[24]和表征^[25]）、机器人技术（例如试剂制备和不同实验模块之间的样品转移）和 AI 算法、数据库融合为一个研发 workflow。通过 AI 算法，系统能够根据历史数据预测最优实验方案，利用机器人技术自动执行样品制备和表征分析，再基于获得的数据实时地调整实验方案，并闭环反馈给实验环节，从而高效地优化目标性能^[26~28]，在几乎不依赖人工干预的情况下进行实验设计、执行、数据收集和分析，自主完成快速迭代。这种智能化的实验方法已在纳米材料合成^[20,29]、催化剂开发^[30]以及功能薄膜制备^[31]等研发任务中展示出巨大的潜力，也为解决复杂科学问题提供了新的可能性。近期，《自然》杂志将其列入 2025 年值得关注的七项技术之一^[21]，可预见其将成为新材料研发范式变革的“新基建”。

高质量的大量数据是 AI 的基础。作为一种基于 AI 的高效材料研究方法与策略，SDL 使用并产生大量的数据，其精髓在于保证数据在工作流程中实时地顺畅流动，采用数据标准极其重要。机器人在操作重复性任务时比实验员更具优势，可看成是大批量、系统性产生标准化材料数据的数据工厂^[32]。目前，采用公共材料数据标准的问题尚未受到广泛关注，造成了自产数据的使用局限于自驱动实验系统内部的闭环反馈流程，难以与系统外进行流通共享，不能充分发挥其价值。因此，数据标准化成为自主实验室从“孤立智能”迈向“协同智能”的关键瓶颈。

本文将综述 AI 驱动的自主实验系统在材料科学中的最新进展，重点关注数据采集与处理、AI 算法创新、实验自动化以及数据标准化等方面的关键突破，并探讨未来发展的挑战与机遇。随着数据标准化技术的不断进步，我们有理由相信，AI 将更加深入地融入实验科学研究的各个环节，推动人类对自然规律的理解迈向新的高度。

1. 人工智能驱动自主实验的技术进展

1.1 自主实验系统与关键技术

自驱动实验系统（SDL）亦称为自主实验系统（autonomous laboratory），指一种高度自动化的科研平台，通过整合机器人技术、AI 算法和自动化实验流程，能够在不依赖人工干预的情况下，完成从实

验设计到结果分析的闭环全过程^[33]，尤其适用于材料科学、化学反应等复杂过程的优化^[34,35]。它与传统的自动化系统（automatic system）的区别在于能形成一个闭环反馈控制系统，不仅能执行自动化的实验操作，而且能对结果进行分析、自主优化实验设计并反馈进行下一轮实验^[36]。SDL 的核心是 AI 算法、数据和机器人技术的深度融合，人工智能算法负责分析实验数据^[37~42]、预测可能的结果^[43]，并设计下一步实验流程^[44]；而机器人则负责执行这些实验，从物料准备到材料制备、样品传递、测试表征，实现全过程的自动化^[45]。该系统通常包括以下几个关键组成部分（图 1(a)）：

- (1) **实验设计模块（experiment design module）**：负责根据研究目标设定具体的实验任务（如寻找具有特定光电性能的材料），以及需要探索的变量（如温度、压力、前驱体比例等）。初始阶段通常涉及已有文献、知识的搜集（计算和实验的）和分析处理，形成决策。后期则依据前期实验数据与优化结果。
- (2) **材料制备模块（materials preparation module）**：基于实验任务定义，该模块利用相应的手段进行材料合成，包括固态、液态或气态的合成方法，样品形态可以是固体、粉末或液体，供性能表征模块使用。
- (3) **性能表征模块（property characterization module）**：此模块对材料制备模块合成的样品，使用分析仪器对性能、结构和微观组织等进行表征，形成所研究样品的构效关系。
- (4) **数据分析模块（data analysis and optimization module）**：该模块对制备和表征数据进行采集、分析，并根据结果反馈调整后续的实验方案。例如，利用深度学习（deep learning, DL）^[46,47]算法对材料的结构与性能关系进行建模，从而预测出更优的合成条件。

Volk 等人^[48]从自主度、操作寿命、实验通量、实验精度、材料用量、可访问参数空间和优化效率等七方面提出了衡量自驱动实验的性能指标。Gary Tom 等人^[49]将自主实验系统按实验工作流程自动化（硬件）和数据驱动决策自主化（软件）两个维度进行分类，参考汽车自动驾驶中的定义，组合得到自主实验系统的等级 0 至等级 5 六个等级（图 1(b)），在硬件层面主要关注系统自动化的程度，可划分为手工实验、单次实验自动化、流程自动化以及全自动化四个层次，其中需要依赖机器人技术^[50]来衔接从样品制备到表征测试的全过程。目前在 SDL 中能进行自动化分析测试的样品以粉末、薄膜和软物质为主，已嵌入 SDL 系统的自动化分析测试技术以谱学表征^[51]为主，包括紫外-可见光谱（ultraviolet-visible spectroscopy, UV-Vis spectroscopy）^[20,25,45,52]、光致发光光谱（photoluminescence spectroscopy, PL）^[20,24,25,52~54]、动态光散射（dynamic light scattering, DLS）^[22]、电泳光散射（electrophoretic light scattering, ELS）

^[22]、X 射线衍射 (X-ray diffraction, XRD) ^[15,51,53,55]、X 射线荧光 (X-ray fluorescence, XRF) ^[53]等, 还有力学^[56]、电学^[57]等性能测试。由于自动化表征与样品合成设备被嵌入至同一自主实验系统后可实时接收新合成样品和产生谱学数据, 它们亦被称为实时/原位谱学。另一方面, 对于块体样品, 目前关于使用自动化 XRD 等谱学表征分析的研究报道较少, 这是未来需要突破的一个局限性。

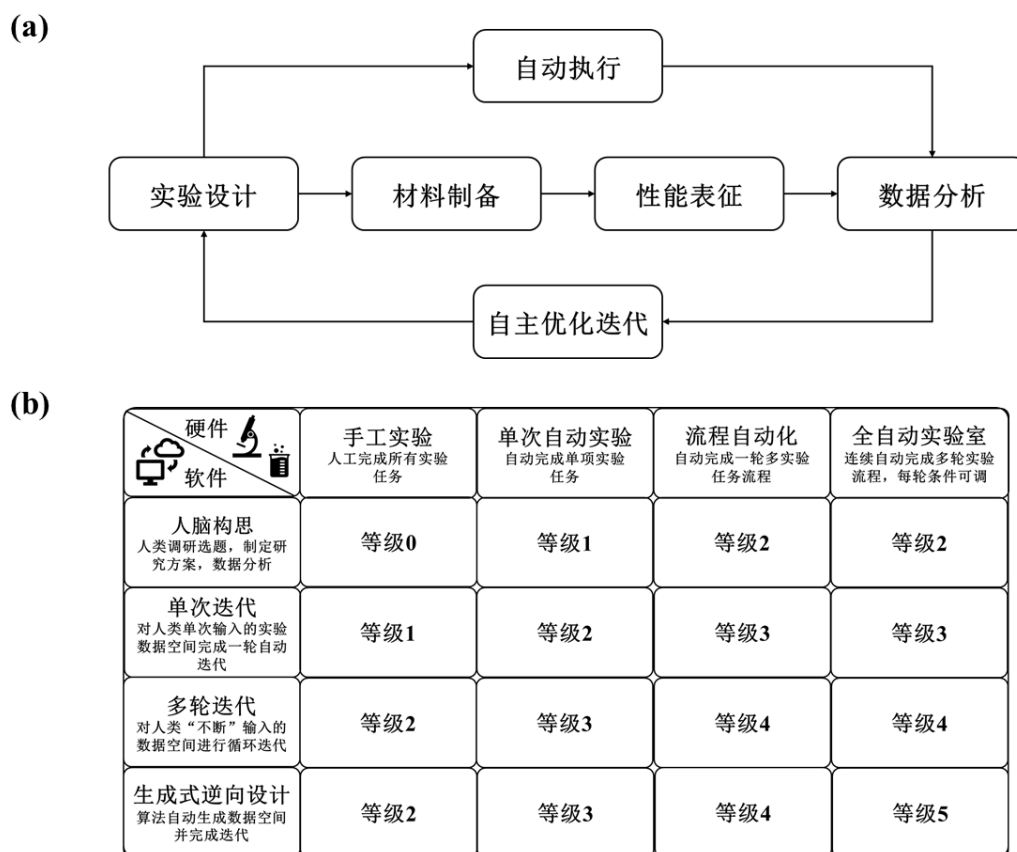


图 1 自主实验系统的(a)构成及(b)自动化或智能化程度的分级 (根据文献[49]重新绘制)

Figure 1 The (a) components and (b) levels of automation or intelligence of the autonomous laboratory (modified from Ref. [49])

软件控制是自主化实验系统的中枢, 区分了实验系统能实现决策自主化的层次, 可划分为人脑构思、单次迭代、多次迭代以及生成式逆向设计四个层次, 其中的关键技术包括: 贝叶斯优化 (Bayesian optimization) ^[58,59]、多目标优化 (multi-objective optimization) ^[60]、知识集成 (knowledge integration)、强化学习 (reinforcement learning) ^[45]等。利用 AI 算法优化分析并反馈新的实验方案是自主实验系统体现其“自主性”的最核心功能, 强化学习、贝叶斯优化等 AI 算法^[43,61~63]可根据自主实验系统硬件端产生的数据对实验方案直接优化, 也可基于特定科学问题先对预测参量做出优化再来解析科学问题^[39,64~66]。

最新的研究指出，对 SDL 实时反馈的高维实验数据采用强化学习有可能比传统的贝叶斯优化更高效地处理高维变量空间的设计优化问题^[67,68]。随着生成式人工智能技术的发展完善，开始涌现出以微软的 MattGen 为代表的生成式材料逆向设计技术^[69]，通过学习已有的知识（包括计算和实验），具备了一定的根据特定性能要求预测可能的化合物和结构的能力，在已有的高通量筛选技术的基础上，具备在更广的成分和结构空间中预测新材料的能力，从而指挥 SDL 进行实验研究。可以预见，在不远的将来，将涌现出这种等级 5 层次的 SDL 系统。

1.2 国内外典型的自主实验案例

尽管自 20 世纪七八十年代起报道了一些自主闭环工作流程的案例，但直到 21 世纪实验室机器人技术与更先进的机器学习算法的集成，才使智能化自主实验系统得到长足发展。Krishnadasan 等人^[70,71]根据在线光谱仪的数据，通过优化自定义实用功能，展示了一种闭环的微流控自主实验系统，能够合成尺寸可控的硒化镉纳米颗粒。英国利物浦大学 Cooper 团队开发了一个名为“移动机器人化学家（mobile robotic chemist）”的自主实验装置^[59]，采用 Kuka 公司的机器人搭建了集成合成和分析功能的自动实验平台，具备利用 AI 算法自主决定下一步化学实验的能力。借助人工智能，移动机器人化学家可以现场更快地做出与研究人员同样的决策，并不间断地自动执行。Coley 等人^[72]报道了一个有机化学智能材料合成系统，结合了人工智能规划和机器人自动执行，可通过分析数百万已发表的反应来设计合成路线，实现了有机化合物的可扩展、可重复合成。中国科学院深圳先进技术研究院的赵海涛等人^[29]报道了一个用于合成胶体纳米晶体的机器人平台，平台通过数据挖掘文献中的初始关键合成参数，结合机器学习模型，实现了对纳米晶体形态的控制合成。

表 1 列出了目前世界上主流材料研究团队的智能化或自主实验系统及各自的特点，以下总结了一些有代表性的系统。

(1) 英国利物浦大学的机器人化学家

英国利物浦大学研究人员开发了一个名为移动机器人化学家（mobile robotic chemist）的自主实验装置^[34,59,73,74]，用于探索和优化光催化剂配方。采用 Kuka 公司的机器人搭建了集成合成和分析功能的自动实验平台，特别适合于在进行光敏感反应研究的全黑环境中操作。硬件上，该系统集成了 6 个模块化工作站，通过 Kuka 机器人能够实现从试剂准备、反应混合到产物分离和表征的全流程自动化操作；制备方法上，该实验装置主要使用纳米粉末颗粒的固态和液态分散法；分析技术上，分别使用光解站和气相色谱仪来评价样品的光催化性能；软件方面，主要使用了机器学习算法（贝叶斯优化）分析实

验数据并预测最优实验条件，从而提高实验效率和成功率，并且系统能够根据实验结果不断调整策略，记录所有实验细节以供后续分析和优化。由于集成了机器人的自动化能力和软件优化决策能力，该系统具备快速迭代和优化的高通量实验能力，能够在短时间内完成大量实验组合，加速新材料的发现过程。在一个全局搜索有接近 1 亿种可能的组分空间中，开展了八天的自主工作，进行了 688 次实验，筛选了 30 种候选基团、测试了不同的染料敏化效果、调整了 pH 值和离子强度的影响，以及评估了表面活性剂和氢键锚定物质的效果，最终将催化剂活性提高了 6 倍。

(2) 劳伦斯伯克利国家实验室的自主新材料发现合成系统 A-Lab

劳伦斯伯克利国家实验室的研究团队与 Google DeepMind 合作^[55]，开发了一个将机器人技术与 AI 相结合的自主新材料发现合成系统 A-Lab，主要用于新型无机材料的合成与优化。系统包括多个工作站：前驱体准备站、带 4 个箱式炉的高温加热站、用于粉末回收和样品加载的样品处理站，以及配备了粉末 X 射线衍射仪的表征站。A-Lab 的自动化程度非常高，几乎涵盖了从实验规划到执行和分析的整个材料合成过程，包括实验规划、样品制备、加热处理、产物表征、数据分析和迭代优化。在系统的控制决策部分，A-Lab 首先利用 Materials Project 提供的大规模第一性原理相稳定性数据来筛选出需要合成的目标材料；然后采用学习算法基于历史文献数据推荐前驱体和合成条件^[75]，经机器人自动化合成和 X 射线衍射（XRD）表征后^[55]，使用机器学习模型分析结果，若合成未达到预期产率或未能形成目标结构，系统会触发主动学习机制，基于 ARROWS（Autonomous Reaction Route Optimization with Solid-State Synthesis）算法^[55,76]提出改进配方和合成条件，形成闭环。在 17 天的连续实验中，自主完成了 355 次实验，合成了 58 个目标化合物中的 41 个，成功率达到 71%^[55]。

(3) 美国马里兰大学的自主材料搜索引擎

组合材料芯片技术可以在一块芯片上制备成千上万不同成分的区域^[8,77]。针对金属材料的相图问题，马里兰大学团队开发了一套自主材料搜索引擎（Autonomous MAterials Search Engine, AMASE）^[4,5,33,78~85]，使用高通量薄膜沉积技术，制备了梯度成分的薄膜样品，通过实时实验-理论计算的闭环互动来自主进行相图的探索与验证，尤其是提高了复杂相图的预测精度及加速了发现新材料的过程（图 2）。AMASE 的核心包括薄膜沉积制备平台、热处理、XRD 表征（在实验室或同步辐射光源完成）、其他物性测量（如椭圆偏振）以及 CALPHAD 热力学计算模型等若干组成部分，通过贝叶斯优化算法进行协调指挥，使得系统能够动态地调整实验成分点，从而构建了实验-理论实时交互的闭环 workflow。系统的数据分析核心是基于自行开发的卷积神经网络的一维 YOLO（You Only Look Once）算法从 XRD 图谱中快速准确地识

别和定位所有主要衍射峰，从而确定样品成分点处的相组成。该系统使用组合材料薄膜共溅射制备技术，可一次性覆盖一个完整的二元或宽阔的多元成分空间体系，特别适用于测定金属材料的相图。在软件控制方面，基于理论预测与实验验证的无缝集成，通过贝叶斯优化算法自主选择下一个最有价值的成分点进行实验表征，大幅减小了实验工作量。该团队以 Sn-Bi 二元系共晶相图为例，展示了该系统在使用成分梯度的薄膜样品快速绘制温度-成分相图方面的应用。AMASE 在 8.4 小时内完成相图测定，较传统网格法效率提升 7 倍，且通过蒙特卡洛采样量化相边界预测不确定度。该系统的局限在于仅支持单一成分梯度样品分析和优化，尚未实现跨体系自主制备及优化。

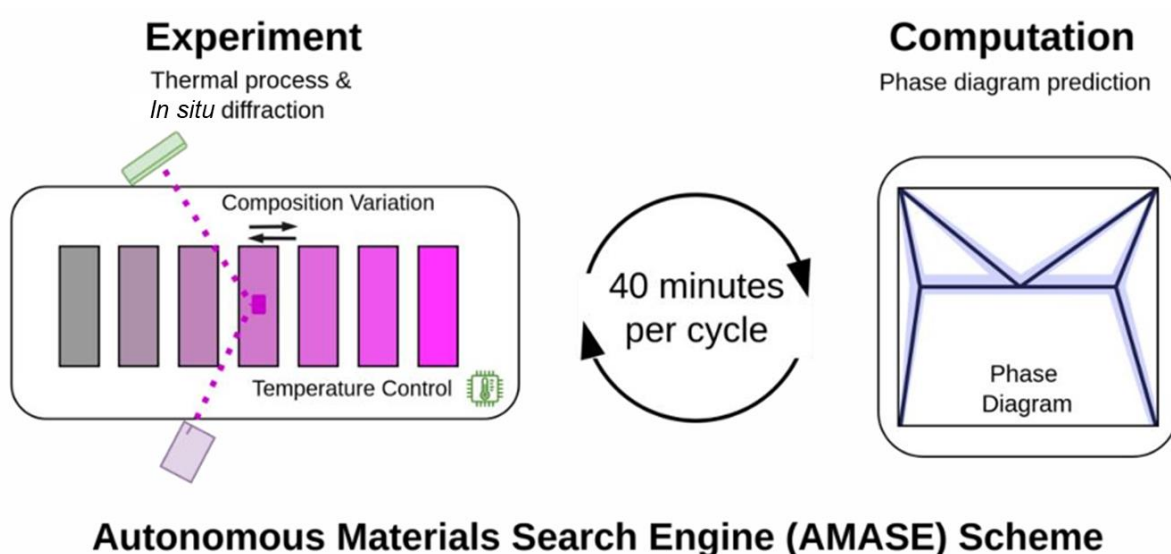


图2 (网络版彩色) 自主材料搜索引擎体系^[33]

Figure 2 (Color online) The autonomous materials research engine (AMASE) system^[33]

(4) 中国科学技术大学的 AI 化学家平台

中国科学技术大学团队^[28,51,86-88]通过开发和集成移动机器人、化学工作站、智能操作系统、科学数据库，研制出数据智能驱动的全流程机器化学家 (AI 化学家) 系统，实现了大数据与智能模型双驱动下的化学合成-表征-测试全流程^[89,90]，可应用于光催化与电催化材料、发光分子、光学薄膜等材料开发。该系统能够根据提出的科学问题自动生成假设和实验计划，并执行完整的实验流程，包括合成、表征和性能测试并分析数据，形成一个完整的闭环工作流程，是一种集合了数据挖掘、自动化合成、机器学习模型及逆向设计的自主实验系统，旨在替代传统试错合成和劳动密集型表征的研究模式^[29]。例如开发高熵

化合物催化剂^[88]，人工仅限于对最多 3 种金属组合进行优化，而 AI 化学家发挥其数据驱动和智能优化的优势，智能阅读 1.6 万篇论文并采用 IUPAC 命名法和 SMILES 分子描述标准，将论文中的合成路径转化为可执行的机器人指令，减少人工干预 80%；自主遴选出 5 种非贵金属元素，融合 2 万组理论计算数据和 207 组全流程机器实验数据，建立了理实交融的智能模型，指导贝叶斯优化程序从 55 万种可能的金属配比中找出最优的高熵催化剂^[28]，将传统“炒菜式”遍历搜索所需的 1400 年缩短为 5 周。与报道的其他自主实验系统相比，其主要特征为：1) 全面性：不仅能读取和理解现有化学知识，还能够设计和执行实验，并基于实验结果提出新的假设；2) 智能化：结合了自然语言处理、机器学习和贝叶斯优化等多种先进技术，使得系统具有自我学习和优化的能力；3) 可扩展性：该系统具有通用的软件协议和标准化的硬件接口，模块化设计使其可通过添加更多的实验工作站或计算机来扩展，以满足各种实验任务的需求。

(5) 远程多地自驱动实验室

上述案例表明，将自动化实验和 AI 决策整合形成的 SDL 工作流，有能力大大加速较为具体的任务目标的实现，例如优化反应条件或确定理想的加工或配方参数。然而，当代前沿材料的发现有时需要更加复杂的合成、配方和表征序列，并配备特定的专业特长或仪器，这些序列的地理分布可能跨越多个地区甚至国家。加拿大、美国、英国、日本、波兰、西班牙等多国的科学家们展示了一种基于云端协同的开发用于有机固体激光器的有机小分子增益材料开发的系统（图 3）^[91]。除了在本地部署闭环的自驱动实验装置外，该系统结合了多个实验室的资源和技术能力，实现了跨时区、跨地点的协同工作，其主要构成包括：1) 中央云服务器：作为整个系统的“大脑”，负责数据管理和分析，并使用机器学习和贝叶斯优化算法对实验进行优先级排序和优化；2) 自动化合成与表征模块：包括自动化的化学合成工作站和光谱表征设备（如 UV-Vis 光谱仪），能够在不同实验室中并行运行；以及 3) 远程协作平台：支持不同实验室之间的实时数据共享和沟通，确保信息流畅传递和协同工作。该系统从大规模分子库中筛选出具有潜在应用价值的有机激光材料^[91]，包括分子设计与合成，利用迭代 Suzuki-Miyaura 偶联反应方法，快速合成大量候选化合物；通过紫外可见光谱、荧光光谱等手段，评估候选材料的光学性能；以及基于代理目标如增益截面进行多轮优化，最终在两个月的优化活动中，发现了 21 种具有改进发射增益截面的小分子发射体，并对选定的最优候选材料在器件上进行了详细性能测试，确认最优候选材料和实际应用潜力^[24]。这种云端协同^[85,91]的远程多地自驱动实验室系统，为智能化跨地区联合攻关的工作模式树立了榜样，展示出在面对更加复杂的材料目标时，下一代人工智能驱动实验可能演变为在地理分布和离域条件下进

行。而这种世界范围的跨实验室、跨地区协同，对数据标准的统一提出了进一步要求，从而保证数据在更大的地域范围内顺畅地流通，通过数据的本地+远程多重循环机制，实现从局部到整体多层次模型优化的大闭环反馈优化过程。

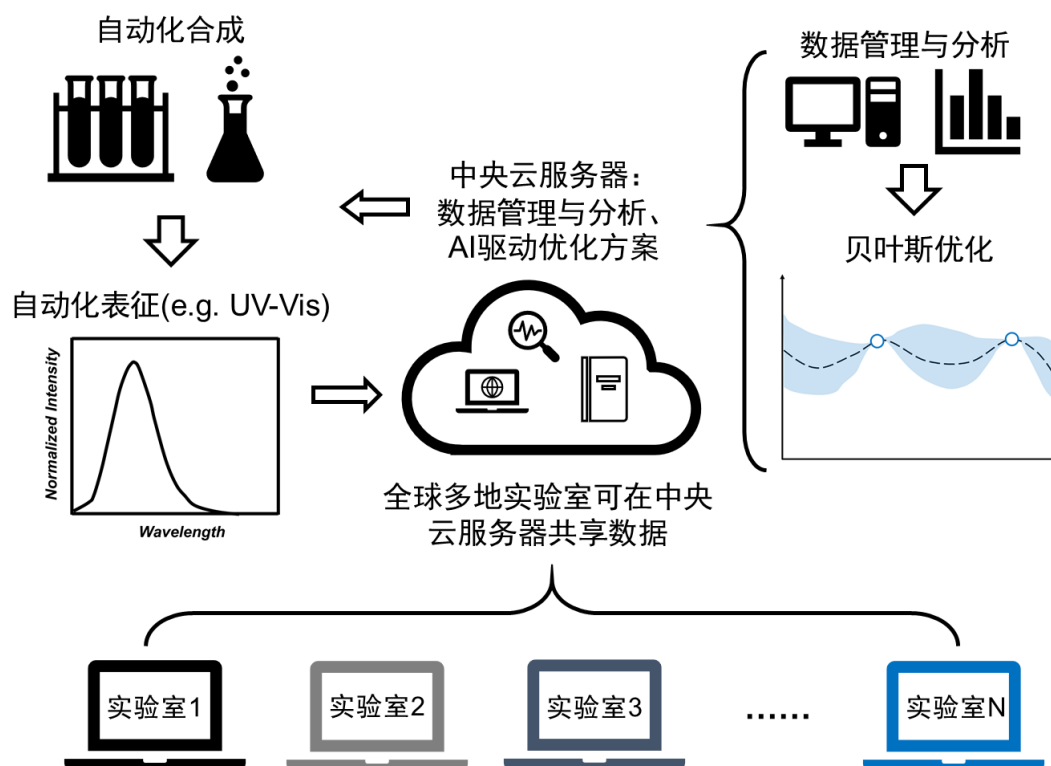


图3 (网络版彩色) 远程多地分散式协同的加速开发有机固体激光器小分子增益材料的工作流程概览^[91]

Figure 3 (Color online) Schematic workflow of the remote dispersive collaborative acceleration of the small molecule gain materials development of the organic solid-state lasers^[91]

表 1 目前世界上主流材料研究团队的智能化或自主实验系统及各自的特点

Table 1 The current intelligent or autonomous laboratories from the leading teams around the world and their respective features

序号	机构	实验技术				自主化程度				应用			
		制备技术	前驱体形态	产物形态	表征技术	自动化实验	机器人	智能化预测	自动化反馈	全流程智能化	材料体系	材料应用方向	优化参数
1	美国麻省理工学院 [72]	流动化学	液态反应物	固态或液态有机物	未明确提出,可能包括色谱、光谱等	部分	固定	否	否	否	有机化合物,包括药物分子	有机合成,药物制造	反应条件 (如温度、压力、流量等)
2	美国马里兰大学 [33]	共溅射制备薄膜	靶材	亚微米薄膜	XRD	是	否	是	是	否	Sn-Bi 二元体系	薄膜器件制备,如超导 $\text{Sn}_x\text{Bi}_{1-x}$ 薄膜	相边界确定、吉布斯自由能最小化
3	中国深圳人工智能与机器人研究院 (AIRS) [92]	微流控反应器、热注入法、机器人自动化控制	溶液	纳米晶体	TEM/HRT、EM、XRD、圆二色光谱 (CD)、吸收光谱、飞秒瞬态吸收光谱	是	是	是	是	否	无机钙钛矿 (CsPbBr_3)	生物传感、手性催化、手性光子学	反应温度、前驱体浓度

4	英国利物浦大学 ^[73]	化学合成 (结构多样化、超分子主客体化学、光化学合成等)	液体、粉体	液体、固体	GC、UPLC-MS、NMR、光催化、XRD	是	移动	是	是	是	光催化材料, 有机合成、超分子化学、光化学	生物衍生空穴清除剂探索, 药物合成、超分子化学	多组分混合物筛选与优化, 化学反应选择性、产物结构与性能等
5	中国科学技术大学 ^[28]	自动合成、自动表征、自动性能测试	液体、粉体	液体、固体	紫外-可见-近红外吸收光谱、荧光光谱、拉曼光谱、电化学工作站、气相色谱等	是	移动	是	是	是	多种无机材料(如氧化物、磷酸盐等)	电催化、光催化、荧光材料等	前驱体选择、反应温度、合成路径
6	中国科学院深圳高等研究院 ^[29]	自动合成、自动表征	液体	液体	紫外-可见-近红外吸收光谱、RGB 值、透射电子显微镜、扫描电子显微镜	部分	移动	是	否	否	金纳米晶体、双钙钛矿纳米晶体	光学、光化学、电化学、光电、生物医学等	结构导向剂(SDAs)含量

7	美国加州伯克利大学, 劳伦斯伯克利国家实验室 [55,93]	固态合成	固体粉末	固体粉末	XRD	是	是	是	是	是	多种无机材料(如氧化物、磷酸盐等)	材料合成验证	前驱体选择、反应温度、合成路径
8	加拿大不列颠哥伦比亚大学 [24,91]	燃烧合成、旋涂、热退火	溶液	薄膜	X 射线荧光光谱 (XRF)、四探针电导测量、暗场摄影、紫外 - 可见 - 近红外反射和透射光谱、四探针电导钨	是	固定	是	是	是	(Pd) 薄膜、有机空穴传输材料 (spiro-O MeTAD)	电子催化、柔性电子、钙钛矿太阳能电池	燃烧合成条件、掺杂比、退火时间

9	埃尔朗根- 纽伦堡弗 里德里希- 亚历山大 大学 [17,58,94]	自动混合、 滴涂、自动 旋涂、喷墨 打印、高温 退火	溶液(有机 体系), 粉 末	薄膜、器件	电 流 密 度 电压 (J - V) 测试、紫外 -可见吸收 光谱、光致 发光 (PL) TRPL 、 XRD 、 SEM、电化 学 表 征 (CV)、局 部 电 化 学 扫描流动 池 (SFC)	部分	固定	否, 是	否, 是	是否	有 机 太 阳 能 电 池 (OSCs), 钙钛矿、有 机半导体、 无 铅 钙 钛 矿	光 伏 光 电 器 件	分子结构, 工艺参数, 供体/受体 组合、有效 带隙 $E_{g,eff}$ 、 形貌, 阳离 子 比 例 (MA/Cs/ Rb/K)、过 化 学 计 量 比、光伏光 稳定性
---	--	--	----------------------	-------	---	----	----	------	------	----	---	----------------	---

2. 数据标准规范在自主实验室发展中的作用

数据是人工智能的基础，人工智能需要大量与之适配的（AI-Ready）数据来支撑^[32]。SDL既是数据的使用者，同时也是数据的生产者。自主实验的初始条件确定，往往需要使用来自实验、计算、文献的既有数据与知识，在实验中系统将实时产出实验结果和数据，用于优化迭代。作为人工智能与自动化技术深度融合的科研范式，其高效的数据产生方式一方面可视为智能化数据工厂^[32]的一种形式，自动化、不间断、流水线式地标准化批量生产材料数据；另一方面，其产生的数据如能与外界进行交换，参与流通，为全社会共享，则可在科学研究协同层面上形成大循环，这种流通共享的前提就是数据标准化。因此，数据标准化是自主化实验室的“通用语言”与智能基石。

目前有关 SDL 的文献中对数据标准讨论较少，由此可以推断出数据在特定系统中的采集与使用依照自有的一套数据标准，能够实现系统内部的数据流通。然而，如果 SDL 产生的数据不采用公共认同的统一数据标准，则仅限于自身内部流动，无法与外界进行流通共享，容易形成“数据孤岛”。这使得 SDL 的作用限制在局部材料开发自循环优化，是一种“孤立智能”。而其作为实验数据工厂的功能无法顺利体现，难以构成全社会数据的大规模流通共享，实现我国在《新材料大数据中心总体建设方案》中提出的，通过“材料+数据”助力新材料原始创新，聚焦汇聚共享资源，搭建新材料大数据中心总体架构，着力打造资源富集、高效贯通、应用繁荣、治理有序的新材料大数据生态的发展目标，从而充分兑现“协同智能”的潜力，这将是一件令人遗憾的事情。

随着 SDL 在材料研发中的普及，缺乏统一的数据标准已成为制约 AI 效能发挥的瓶颈。数据标准化是 SDL 从“孤立智能”迈向“协同智能”的关键瓶颈，短期内需优先解决元数据规范和跨平台兼容性问题，长期依赖全球科研生态的协作重构。对于跨平台兼容性问题，国际上，欧洲的材料设计开放数据库集成联盟（Open Databases Integration for Materials Design, OPTIMADE; <https://www.optimade.org/>）发布了一种统一的应用程序编程接口（API）标准，通过 OPTIMADE API 可以让物理位置分布不同的材料数据基础设施实现跨库数据检索，例如美国劳伦斯伯克利国家实验室的 A-Lab 与多伦多大学的 ORGAN 系统能直接调用彼此数据库中的 5.6 万种钙钛矿数据，将新型光伏材料发现效率提升 4 倍^[17,20,29,92,94,95]。对元数据规范问题，欧盟的新材料发现卓越中心（Novel Materials Discovery, NOMAD; <https://nomad-lab.eu/nomad-lab/>）开发了材料科学领域迄今最完备的元数据框架，该框架已被 ISO/TC 329 采纳为国际标准草案，全球超过 200 个材料数据库实现兼容。NOMAD 的材料数据管理和分享框架中已将实验数据列为其重要的组成部分。我

国率先于2019年发布了由国内30余家材料研究主体单位共同制定的世界范围内首个关于材料基因工程数据的团体标准——《材料基因工程数据通则》（T/CSTM 00120-2019, <https://ndls.org.cn/standard/detail/0122114f8887f98053ea4fe1dd7c37ab>），对AI-Ready数据的内容进行了原则上的规定。随后又相继发布了基于《材料基因工程数据通则》为原则的《材料基因工程技术语》（T/CSTM 00839-2022, <https://ndls.org.cn/standard/detail/c4595a3e5b0fd8c7d60662693889e1a4>）、《材料基因工程数据元数据标准化原则与方法》（T/CSTM 00837-2022, <https://ndls.org.cn/standard/detail/90eb9beae28efea409da52ae26c3d4e3>）、《材料基因工程材料数据标识（MID）》（T/CSTM 00838-2022, <https://ndls.org.cn/standard/detail/1b401dae3f7ea539f325480bee47d809>）等12项相关标准，系统性建设材料数据标准体系的工作正在进行中。

随着对数据的认识不断深化，材料数据标准正从辅助工具演变为自主化实验室的重要基础设施。这里，我们提出“标准植入的智能化系统”的概念模型，以材料标准化大数据管理平台、人工智能模块和高通量数据生产系统为核心（如图4所示），全面植入AI-Ready数据标准，构建出“高通量数据生产-标准化数据库-人工智能”的数据闭环内/外双循环自主化实验模式。内循环的环路包括一套单一自主实验系统，通过高通量实验与高通量计算方法实现数据的批量化生产，材料标准化大数据管理平台依据统一的数据标准采集、管理数据，人工智能模块基于标准化的材料数据集进行模型训练与决策，实时优化高通量数据生产模块的方案设计与参数迭代。与此同时，材料标准化大数据管理平台通过标准化接口将数据传输至国家材料大数据主平台，并通过主平台于新材料大数据中心各个节点进行实时数据流通与共享；当所有SDL都以同样方式连接到国家新材料大数据主平台，我们将看到一个服务于人工智能的、为全社会所共享的新型材料数据基础设施与标准化数据生态的形成，数据作为一种社会资源在两个层面进行着循环，为实现高效的数据驱动研发奠定了设施基础。

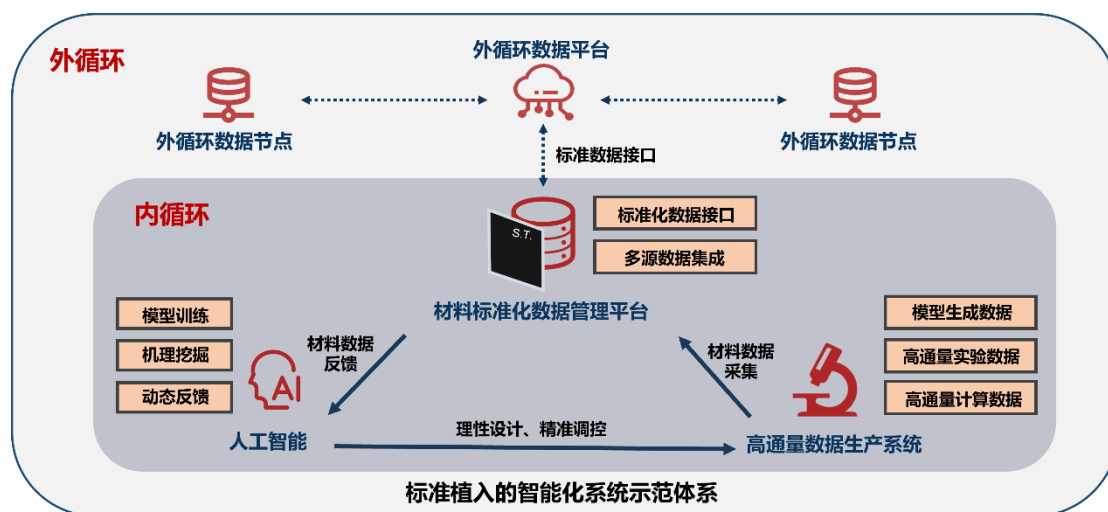


图4 （网络版彩色）高通量数据生产-标准化数据库-人工智能的数据闭环内/外循环模式

Figure 4 (Color online) The internal/external data cyclic modes of the “production of high-throughput experimental data-standardized database-artificial intelligence”

数据的规模^[96~99]和数据的质量^[100~102]是影响 AI 模型质量的一个重要因素，充足、高质量的领域科学数据是人工智能参与科学研究的关键。与来源广泛的互联网数据相比，领域科学数据具有更强的专业性，获取成本也更高，一般由具备特定领域专业知识的人员来产生、整理和使用。2023 年，Liu 等人^[103]提出数据数量治理应关注样本数量与特征数量（如描述符）以及二者指标的平衡性，进而提出融合材料领域知识的数据数量协同治理流程，实现了样本规模与特征维度（或模型参数量）的精准治理与协调。同年，刘悦等人^[104]又进一步提出材料领域知识全过程嵌入的材料机器学习数据质量与数量治理框架，实现了材料数据质量和数量的整体评估与提升，为高质量和适度数量的数据获取提供了理论指导和候选解决方案，推动机器学习在材料研发中的深入应用。

3. 挑战与未来发展方向

尽管智能化/自主化系统革新了传统材料研发范式，其技术边界仍在许多方面制约着跨领域协同创新。例如：

适用于智能化系统的材料制备方法较为局限：现有系统多聚焦溶液和粉末化学合成（如机器人流动化学平台的液态前驱体处理^[72]、A-Lab 的固态粉末反应^[55]），而对物理沉积（如磁控溅射、分子束外延）等复杂工艺的兼容性不足。以薄膜材料为例，尽管 Du 等人^[105]用 AMANDA Line One

自主实验平台实现了有机太阳能电池材料的旋涂自动化,但对物理气相沉积等涉及超高真空与界面工程的工艺仍缺乏适应性,导致其在宽禁带半导体、超晶格结构等前沿领域的应用仍然空缺^[24]。

系统通用性与迁移性缺失成为瓶颈:现有平台高度依赖预设材料体系(如纳米晶体机器人平台专攻纳米晶体^[48], KUKA 移动平台聚焦光催化制氢^[59]),其模块化设计往往难以适配跨相态(固-液-气)、跨尺度(微观-介观-宏观)的合成需求。因此,需要开发多种丰富的制备合成手段和适配材料性能表征手段,例如:除了湿法合成、薄膜合成手段外,需要开发合适自动化的块体合成手段以满足大量的块体材料研究的需求;同时,从构建材料构效关系的角度出发,开发更多适配于自动化操作的分析表征手段如电子显微镜、物理性能测试、结构性能测试等,进一步扩大自驱动实验室的研究领域。

自主决策能力的局部性问题凸显——当前多数智能化系统仅在实验执行层实现自动化(如纳米晶体机器人平台^[23,48], OSL 自主实验室^[24]等),而在跨模态数据融合、非线性优化反馈等环节仍依赖人工经验,需要研究者/工程师介入不同实验步骤的传递或者反应路径的迭代决策,阻碍了全闭环智能的实现。随着 AI 技术的不断进步,未来的研究应致力于开发更加高效和智能的 AI 算法,以进一步优化实验设计流程。通过机器学习模型的深度学习能力,可以实现对复杂材料体系的预测与优化,从而提高实验效率并降低资源消耗^[106]。此外,结合强化学习^[23,45]等技术, AI 可以在动态调整实验参数的过程中快速迭代,推动材料性能的突破性提升。研究者需要进一步探索多目标优化算法,以在材料性能、成本和环境友好性之间找到最佳平衡点。

现代化的 SDL 是高度定制化的,因此结构复杂、价格昂贵。典型的 SDL 设备成本普遍超过百万美元级别,设备专用性强,限制了其推广应用,也考验研究团队的资金筹措能力。为了推动自驱动实验室的大规模普及,加拿大多伦多大学团队^[107]提出了“节俭孪生”(frugal twin)策略,采用开发通用的模块化的自动化硬件,统一接口和用户界面,令自驱动实验室的搭建可以用类似乐高积木的方式构建。使用的硬件模块在系统开发建设过程中可先选择低成本产品,完成系统搭建;在实际使用时再根据实验实际需求选择替换,丰俭由人。通过硬件替代策略、模块化架构设计、开源生态构建等创新,将系统建立的最低成本压缩至千美元数量级,大幅度降低了系统建设成本,将自驱动实验室由专家设计和操作的系统过渡到日常工具,同时也锻炼了队伍,培养了人才。

~~最后~~在推动数据交换的标准化和数据标准的国际化,形成全球、社会协同的大格局的过程中,数据是 AI 驱动实验室的核心资源,数据标准是 SDL 的通用语言与智能基石。然而,当前材料科学领域的数据格式和标准尚未完全统一,导致数据孤岛现象严重,当前材料数据标准在 SDL 中的应用还面临一些挑战;如数据标准的制定和推广需要多方协作,数据交换机制、数据隐私和

安全等关键问题需要解决等。未来的研究应致力于推动数据标准的国际化与标准化，建立标准化协议和数据格式，建立统一的数据共享平台，有效实施 AI-ready 的数据可发现、可获取、可交互、可再利用（FAIR）原则，促进跨机构、跨国界的高效协作。通过制定通用的数据描述规范和元数据标准，可以提升科学的透明度、实验数据的可重复性和可靠性，为全球科学家提供高质量的数据资源支持，形成高效、可信的大规模科研协作。

4. 总结

人工智能驱动的自主实验正在彻底改变材料科学领域的研究方式。传统上，材料科学研究依赖于繁琐的手动实验和试错法，而 AI 的引入使得实验过程更加高效和智能化。通过自动化设备和机器学习算法，系统能够自主完成从实验设计到结果分析的全过程，并根据实时反馈调整实验参数，优化实验条件。这种智能化实验方法不仅提高了科研效率，还为解决复杂科学问题提供了新的可能性。

SDL 既是数据的使用者，同时也是数据的生产者。其高效的数据产生方式可视为智能化数据工厂的一种形式。数据标准化是自主化实验室的“通用语言”与智能基石，当前缺乏统一的数据标准导致了数据孤岛的产生，制约了其数据工厂角色的发挥。尽快推动数据标准的国际化，形成全社会协同的大格局，是 SDL 从“孤立智能”迈向“协同智能”的关键一步。

展望未来，下一代人工智能驱动实验可能演变为在地理分布和离域条件下进行，面向更加复杂的材料目标，开展世界范围的跨地区、跨实验室协同。这种云端协同的远程多地自驱动实验室系统对数据标准的统一提出了进一步要求，通过数据的本地+远程多重循环机制，实现从局部到整体多层次模型优化的大闭环反馈优化过程，与此同时，也将推动服务于人工智能的、为全社会所共享的新型材料数据基础设施与数据生态尽早形成。

参考文献

- [1]. Topham S A. The History of the Catalytic Synthesis of Ammonia. In: Anderson J R, Boudart M, eds. Catalysis: Science and Technology. Berlin: Springer, 1985. 1–50
- [2]. Haber F. The synthesis of ammonia from its elements. Nobel Lecture, June 2, 1920. Reson, 2002, 7(9): 86–94
- [3]. Chen S, Perathoner S, Ampelli C, et al. Chapter 2—Electrochemical Dinitrogen Activation: To

Find a Sustainable Way to Produce Ammonia. In: Albonetti S, Perathoner S, Quadrelli E A, eds. *Studies in Surface Science and Catalysis*. Amsterdam: Elsevier, 2019. 31–46

[4]. Wang A, Liang H, McDannald A, et al. Benchmarking active learning strategies for materials optimization and discovery. *Oxford Open Mater Sci*, 2022, 2(1): itac006

[5]. Adams F, McDannald A, Takeuchi I, et al. Human-in-the-loop for Bayesian autonomous materials phase mapping. *Matter*, 2024, 7(2): 697–709

[6]. Chang M C, Ament S, Amsler M, et al. Probabilistic phase labeling and lattice refinement for autonomous materials research. *npj Comput Mater*, 2025, 11(1): 1–13

[7]. Sanin A, Flowers J K, Piotrowiak T H, et al. Integrating Automated Electrochemistry and High-Throughput Characterization with Machine Learning to Explore Si—Ge—Sn Thin-Film Lithium Battery Anodes. *Adv Energy Mater*, 2025, 15(11): 2404961

[8]. Xiang X D, Sun X, Briceño G, et al. A Combinatorial Approach to Materials Discovery. *Science*, 1995, 268(5218): 1738–1740

[9]. Isenhour T L. Robotics In The Laboratory. *J Chem Inf Comput Sci*, 1985, 25(3): 292–295

[10]. Chavez-Angel E, Eriksen M B, Castro-Alvarez A, et al. Applied Artificial Intelligence in Materials Science and Material Design. *Adv Intell Syst*, 2025: 2400986

[11]. Zhao Q, Zhang Z, Gao H, et al. AI-Powered Visualization of Invisible Mechano-Information: Stress, Defects, and Beyond. *Adv Funct Mater*, 2025: 2506790

[12]. Han X Q, Wang X D, Xu M Y, et al. AI-Driven Inverse Design of Materials: Past, Present, and Future. *Chin Phys Lett*, 2025, 42(2): 027403

[13]. Febba D, Egbo K, Callahan W A, et al. From text to test: AI-generated control software for materials science instruments. *Digital Discovery*, 2025, 4(1): 35–45

[14]. Beaucage P A, Sutherland D R, Martin T B. Automation and Machine Learning for Accelerated Polymer Characterization and Development: Past, Potential, and a Path Forward. *Macromolecules*, 2024, 57(18): 8661–8670

- [15]. Xie Y, Sattari K, Zhang C, et al. Toward autonomous laboratories: Convergence of artificial intelligence and experimental automation. *Prog Mater Sci*, 2023, 132: 101043
- [16]. Cheng D, Sha W, Guo Y, et al. A real-time deep learning model to narrow the gap between atomic scanning transmission electron microscopy and theory calculations: Recognition, reconstruction, and simulation. *MRS Bull*, 2025, 50(2): 101–114
- [17]. Zhang J, Wu J, Stroyuk O, et al. Self-driving AMADAP laboratory: Accelerating the discovery and optimization of emerging perovskite photovoltaics. *MRS Bull*, 2024, 49(12): 1284–1294
- [18]. Zhang J, Hauch J A, Brabec C J. Toward Self-Driven Autonomous Material and Device Acceleration Platforms (AMADAP) for Emerging Photovoltaics Technologies. *Acc Chem Res*, 2024, 57(9): 1434–1445
- [19]. Yotsumoto Y, Nakajima Y, Takamoto R, et al. Autonomous robotic experimentation system for powder X-ray diffraction. *Digital Discovery*, 2024, 3(12): 2523–2532
- [20]. Sadeghi S, Bateni F, Kim T, et al. Autonomous nanomanufacturing of lead-free metal halide perovskite nanocrystals using a self-driving fluidic lab. *Nanoscale*, 2024, 16(2): 580–591
- [21]. Eisenstein M. Self-driving laboratories, advanced immunotherapies and five more technologies to watch in 2025. *Nature*, 2025, 637(8047): 1008–1011
- [22]. Zaki M, Prinz C, Ruehle B. A Self-Driving Lab for Nano- and Advanced Materials Synthesis. *ACS Nano*, 2025, 19(9): 9029–9041
- [23]. Volk A A, Epps R W, Yonemoto D T, et al. AlphaFlow: autonomous discovery and optimization of multi-step chemistry using a self-driven fluidic lab guided by reinforcement learning. *Nat Commun*, 2023, 14(1): 1403
- [24]. Seifrid M, Pollice R, Aguilar-Granda A, et al. Autonomous Chemical Experiments: Challenges and Perspectives on Establishing a Self-Driving Lab. *Acc Chem Res*, 2022, 55(17): 2454–2466
- [25]. Sadeghi S, Mattsson K, Glasheen J, et al. A self-driving fluidic lab for data-driven synthesis of lead-free perovskite nanocrystals. *Digital Discovery*, 2025, doi: 10.1039/d5dd00062a

- [26]. Chen C, Zuo Y, Ye W, et al. A Critical Review of Machine Learning of Energy Materials. *Adv Energy Mater*, 2020, 10(8): 1903242
- [27]. Yang X K, Xu Y Y, Chen L, et al. AI for Science: AI enabled scientific facility transforms fundamental research (in Chinese). *Bulletin of Chinese Academy of Sciences*, 2024, 39(1): 59–69
[杨小康, 许岩岩, 陈露, 等. AI for Science: 智能化科学设施变革基础研究. *中国科学院院刊*, 2024, 39(1): 59–69]
- [28]. Zhu Q, Zhang F, Huang Y, et al. An all-round AI-Chemist with a scientific mind. *Natl Sci Rev*, 2022, 9(10): nwac190
- [29]. Zhao H, Chen W, Huang H, et al. A robotic platform for the synthesis of colloidal nanocrystals. *Nat Synth*, 2023, 2(6): 505–514
- [30]. Xu S, Chen Z, Qin M, et al. Developing new electrocatalysts for oxygen evolution reaction via high throughput experiments and artificial intelligence. *npj Comput Mater*, 2024, 10(1): 194
- [31]. MacLeod B P, Parlane F G L, Morrissey T D, et al. Self-driving laboratory for accelerated discovery of thin-film materials. *Sci Adv*, 2020, 6(20): eaaz8867
- [32]. Wang H, Xiang X D, Zhang L. On the Data-Driven Materials Innovation Infrastructure. *Engineering*, 2020, 6(6): 609–611
- [33]. Liang H, Wang C, Yu H, et al. Real-time experiment-theory closed-loop interaction for autonomous materials science. 2024, arXiv: 2410.17430
- [34]. Lunt A M, Fakhruldeen H, Pizzuto G, et al. Modular, multi-robot integration of laboratories: an autonomous workflow for solid-state chemistry. *Chemical Science*, 2024, 15(7): 2456–2463
- [35]. Delgado-Licona F, Abolhasani M. Research Acceleration in Self-Driving Labs: Technological Roadmap toward Accelerated Materials and Molecular Discovery. *Advanced Intelligent Systems*, 2023, 5(4): 2200331
- [36]. Wang C, Kim Y J, Vriza A, et al. Autonomous platform for solution processing of electronic polymers. *Nat Commun*, 2025, 16(1): 1498
- [37]. Gaines T B, Boyle C, Keller J M, et al. Scanning Electron Microscope Image Segmentation with Foundation AI Vision Model for Nanoparticles in Autonomous Materials Explorations. In:

2nd IEEE Conference on Artificial Intelligence (CAI), 2024. Singapore: IEEE Computer Society, 2024. 1266–1271

[38]. Kaufmann K, Zhu C Y, Rosengarten A S, et al. Crystal symmetry determination in electron diffraction using machine learning. *Science*, 2020, 367(6477): 564–568

[39]. Kalinin S V, Ziatdinov M, Spurgeon S R, et al. Deep learning for electron and scanning probe microscopy: From materials design to atomic fabrication. *MRS Bull*, 2022, 47(9): 931–939

[40]. Genc A, Marlowe J, Jalil A, et al. A versatile machine learning workflow for high-throughput analysis of supported metal catalyst particles. *Ultramicroscopy*, 2025, 271: 114116

[41]. Roccapriore K M, Kalinin S V, Ziatdinov M. Physics Discovery in Nanoplasmonic Systems via Autonomous Experiments in Scanning Transmission Electron Microscopy. *Adv Sci*, 2022, 9(36): 2203422

[42]. Olszta M H, Hopkins D, Fiedler K R, et al. An Automated Scanning Transmission Electron Microscope Guided by Sparse Data Analytics. *Microsc Microanal*, 2022, 28(5): 1611–1621

[43]. Szymanski N J, Bartel C J, Zeng Y, et al. Adaptively driven X-ray diffraction guided by machine learning for autonomous phase identification. *npj Comput Mater*, 2023, 9(1): 31

[44]. Bateni F, Epps R W, Antami K, et al. Autonomous Nanocrystal Doping by Self-Driving Fluidic Micro-Processors. *Adv Intell Syst*, 2022, 4(5): 2200017

[45]. Li J, Tu Y, Liu R, et al. Toward "On-Demand" Materials Synthesis and Scientific Discovery through Intelligent Robots. *Adv Sci*, 2020, 7(7): 1901957

[46]. Szymanski N J, Zeng Y, Huo H Y, et al. Toward autonomous design and synthesis of novel inorganic materials. *Mater Horiz*, 2021, 8(8): 2169–2198

[47]. Rettenberger L, Szymanski N J, Zeng Y, et al. Uncertainty-aware particle segmentation for electron microscopy at varied length scales. *npj Comput Mater*, 2024, 10(1): 124

[48]. Volk A A, Abolhasani M. Performance metrics to unleash the power of self-driving labs in chemistry and materials science. *Nat Commun*, 2024, 15(1): 1378

- [49]. Tom G, Schmid S P, Baird S G, et al. Self-Driving Laboratories for Chemistry and Materials Science. *Chem Rev*, 2024, 124(16): 9633–9732
- [50]. MacLeod B P, Parlane F G L, Brown A K, et al. Flexible automation accelerates materials discovery. *Nat Mater*, 2022, 21(7): 722–726
- [51]. Song T, Luo M, Zhang X L, et al. A Multiagent-Driven Robotic AI Chemist Enabling Autonomous Chemical Research On Demand. *J Am Chem Soc*, 2025, 147(15): 12534–12545
- [52]. Bateni F, Sadeghi S, Orouji N, et al. Smart Dope: A Self-Driving Fluidic Lab for Accelerated Development of Doped Perovskite Quantum Dots. *Adv Energy Mater*, 2024, 14(1): 2302303
- [53]. Stier S P, Kreisbeck C, Ihssen H, et al. Materials Acceleration Platforms (MAPs) Accelerating Materials Research and Development to Meet Urgent Societal Challenges. *Adv Mater*, 2024, 36(45): 2407791
- [54]. Bayley O, Savino E, Slattery A, et al. Autonomous chemistry: Navigating self-driving labs in chemical and material sciences. *Matter*, 2024, 7(7): 2382–2398
- [55]. Szymanski N J, Rendy B, Fei Y, et al. An autonomous laboratory for the accelerated synthesis of novel materials. *Nature*, 2023, 624(7990): 86–91
- [56]. Emery B, Snapp K L, Revier D, et al. Foams with 3D Spatially Programmed Mechanics Enabled by Autonomous Active Learning on Viscous Thread Printing. *Adv Sci*, 2024, 11(44): 2408062
- [57]. Nyayachavadi A, Wang C S, Vriza A, et al. Tunable Solid-State Properties and Anisotropic Charge Mobility in Hydrogen-Bonded Diketopyrrolopyrrole Polymers via Automated Device Fabrication and Characterization. *Adv Funct Mater*, 2024, 34(40): 2403612
- [58]. Langner S, Hase F, Perea J D, et al. Beyond Ternary OPV: High-Throughput Experimentation and Self-Driving Laboratories Optimize Multicomponent Systems. *Adv Mater*, 2020, 32(14): 1907801
- [59]. Burger B, Maffettone P M, Gusev V V, et al. A mobile robotic chemist. *Nature*, 2020, 583(7815): 237–241

- [60]. Vivar M, Fuentes M, Dodd N, et al. First lab-scale experimental results from a hybrid solar water purification and photovoltaic system. *Sol Energy Mater Sol Cells*, 2012, 98: 260–266
- [61]. Maffettone P M, Banko L, Cui P, et al. Crystallography companion agent for high-throughput materials discovery. *Nat Comput Sci*, 2021, 1(4): 290–297
- [62]. Kusne A G, Yu H, Wu C, et al. On-the-fly closed-loop materials discovery via Bayesian active learning. *Nat Commun*, 2020, 11(1): 5966
- [63]. Stein H S, Gregoire J M. Progress and prospects for accelerating materials science with automated and autonomous workflows. *Chem Sci*, 2019, 10(42): 9640–9649
- [64]. Kalinin S V, Ophus C, Voyles P M, et al. Machine learning in scanning transmission electron microscopy. *Nat Rev Methods Primers*, 2022, 2(1): 11
- [65]. Kalinin S V, Ziatdinov M, Hinkle J, et al. Automated and Autonomous Experiments in Electron and Scanning Probe Microscopy. *ACS Nano*, 2021, 15(8): 12604–12627
- [66]. Montoya J H, Grimley C, Aykol M, et al. How the AI-assisted discovery and synthesis of a ternary oxide highlights capability gaps in materials science. *Chem Sci*, 2024, 15(15): 5660–5673
- [67]. Xian Y, Dang P, Tian Y, et al. Compositional design of multicomponent alloys using reinforcement learning. *Acta Mater*, 2024, 274: 120017
- [68]. Xian Y, Ding X, Jiang X, et al. Unlocking the black box beyond Bayesian global optimization for materials design using reinforcement learning. *npj Comput Mater*, 2025, 11(1): 1–11
- [69]. Zeni C, Pinsler R, Zügner D, et al. A generative model for inorganic materials design. *Nature*, 2025, 639(8055): 1–3
- [70]. Krishnadasan S, Brown R J C, Demello A, et al. Intelligent routes to the controlled synthesis of nanoparticles. *Lab Chip*, 2007, 7(11): 1434–1441
- [71]. Krishnadasan S, Tovilla J, Vilar R, et al. On-line analysis of CdSe nanoparticle formation in a continuous flow chip-based microreactor. *J Mater Chem*, 2004, 14(17): 2655–2660

- [72]. Coley C W, Thomas III D A, Lummiss J A M, et al. A robotic platform for flow synthesis of organic compounds informed by AI planning. *Science*, 2019, 365(6453): eaax1566
- [73]. Dai T, Vijayakrishnan S, Szczypinski F T, et al. Autonomous mobile robots for exploratory synthetic chemistry. *Nature*, 2024, 635(8040): 1–8
- [74]. Fakhruddin H, Pizzuto G, Glowacki J, et al. ARChemist: Autonomous Robotic Chemistry System Architecture. In: *Proceedings of the IEEE International Conference on Robotics and Automation*, 2022. Philadelphia: IEEE Robotics and Automation Society, 2022. 6013–6019
- [75]. He T, Huo H, Bartel C J, et al. Precursor recommendation for inorganic synthesis by machine learning materials similarity from scientific literature. *Sci Adv*, 2023, 9(23): eadg8180
- [76]. Szymanski N J, Nevatia P, Bartel C J, et al. Autonomous and dynamic precursor selection for solid-state materials synthesis. *Nat Commun*, 2023, 14(1): 6956
- [77]. Xiang X D, Wang G, Zhang X, et al. Individualized Pixel Synthesis and Characterization of Combinatorial Materials Chips. *Engineering*, 2015, 1(2): 225–233
- [78]. Price C, Takeuchi I, Rondinelli J, et al. AI-Accelerated Electronic Materials Discovery and Development. *Computer*, 2025, 58(2): 98–104
- [79]. Liu Y, Pant R, Takeuchi I, et al. Automated Materials Discovery Platform Realized: Scanning Probe Microscopy of Combinatorial Libraries. 2024, arXiv: 2412.18067
- [80]. Osés C, Esters M, Hicks D, et al. aflow++: A C++ framework for autonomous materials design. *Comput Mater Sci*, 2023, 217: 111889
- [81]. Kusne A G, McDannald A, Takeuchi I. Live Autonomous Beamline Experiments: Physics In the Loop. In: Marcus N, Daniela U, eds. *Methods and Applications of Autonomous Experimentation*. New York: Taylor & Francis Group, 2023. 304–314
- [82]. McDannald A, Frontzek M, Savici A T, et al. On-the-fly autonomous control of neutron diffraction via physics-informed Bayesian active learning. *Appl Phys Rev*, 2022, 9(2): 021408
- [83]. Kusne A G, McDannald A, DeCost B, et al. Physics in the machine: integrating physical knowledge in autonomous phase-mapping. *Front Phys*, 2022, 10: 815863

- [84]. Stanev V, Choudhary K, Kusne A G, et al. Artificial intelligence for search and discovery of quantum materials. *Commun Mater*, 2021, 2(1): 105
- [85]. Oses C, Gossett E, Hicks D, et al. AFLOW-CHULL: cloud-oriented platform for autonomous phase stability analysis. *J Chem Inf Model*, 2018, 58(12): 2477–2490
- [86]. Zhu Q, Huang Y, Zhou D, et al. Automated synthesis of oxygen-producing catalysts from Martian meteorites by a robotic AI chemist. *Nat Synth*, 2024, 3(3): 319–328
- [87]. Zhang B, Zhu Z, Li H, et al. Revolutionizing Chemistry and Material Innovation: An Iterative Theoretical-Experimental Paradigm Leveraged by Robotic AI Chemists. *CCS Chem*, 2025, 7(2): 345–360
- [88]. Luo M, Xie Z, Li H, et al. Physics-informed, dual-objective optimization of high-entropy-alloy nanozymes by a robotic AI chemist. *Matter*, 2025, 8(4): 102009
- [89]. Xiao H, Zhang F, Zhu Q, et al. AI-chemist for chemistry synthesis, property characterization, and performance testing. *Sci Sin Chim*, 2023, 53(1): 9–18
- [90]. Zhou J, Luo M, Chen L, et al. A multi-robot-multi-task scheduling system for autonomous chemistry laboratories. *Digital Discovery*, 2025, 4(3): 636–652
- [91]. Strieth-Kalthoff F, Hao H, Rathore V, et al. Delocalized, asynchronous, closed-loop discovery of organic laser emitters. *Science*, 2024, 384(6697): eadk9227
- [92]. Li J, Li J, Liu R, et al. Autonomous discovery of optically active chiral inorganic perovskite nanocrystals through an intelligent cloud lab. *Nat Commun*, 2020, 11(1): 2046
- [93]. Fei Y, Rendy B, Kumar R, et al. AlabOS: a Python-based reconfigurable workflow management framework for autonomous laboratories. *Digital Discovery*, 2024, 3(11): 2275–2288
- [94]. Zhao Y, Zhang J, Xu Z, et al. Discovery of temperature-induced stability reversal in perovskites using high-throughput robotic learning. *Nat Commun*, 2021, 12(1): 2191
- [95]. Bateni F. Autonomous Microfluidic Synthesis of Metal Cation-Doped Perovskite Quantum Dots. Doctor Dissertation. North Carolina: North Carolina State University, 2025

- [96]. Song Y, Siriwardane E M D, Zhao Y, et al. Computational Discovery of New 2D Materials Using Deep Learning Generative Models. *ACS Appl Mater Interfaces*, 2021, 13(45): 53303–53313
- [97]. Zhao K L, Jin X L, Wang Y Z. Survey on Few-shot Learning (in Chinese). *J Softw*, 2021, 32(2): 349–369 [赵凯琳, 靳小龙, 王元卓. 小样本学习研究综述. *软件学报*, 2021, 32(2): 349–369]
- [98]. Gupta V, Choudhary K, Tavazza F, et al. Cross-property deep transfer learning framework for enhanced predictive analytics on small materials data. *Nat Commun*, 2021, 12(1): 6595
- [99]. Pruksawan S, Lambard G, Samitsu S, et al. Prediction and optimization of epoxy adhesive strength from a small dataset through active learning. *Sci Technol Adv Mater*, 2019, 20(1): 1010–1021
- [100]. Liu Z, Chen C, Zhou Y, et al. Relating the combinatorial materials chip mapping to the glass-forming ability of bulk metallic glasses via diffraction peak width. *Scr Mater*, 2024, 240: 115848
- [101]. Liu Z, Chen C, Zhou Y, et al. An element-wise machine learning strategy to predict glass-forming range of ternary alloys enabled by comprehensive data. *Scr Mater*, 2023, 229: 115347
- [102]. Yao Y, Sullivan T, Yan F, et al. Balancing data for generalizable machine learning to predict glass-forming ability of ternary alloys. *Scr Mater*, 2022, 209: 114366
- [103]. Liu Y, Yang Z, Zou X, et al. Data quantity governance for machine learning in materials science. *Natl Sci Rev*, 2023, 10(7): nwad125
- [104]. Liu Y, Ma S, Yang Z, et al. A Data Quality and Quantity Governance for Machine Learning in Materials Science (in Chinese). *J Chin Ceram Soc*, 2023, 51(2): 427–437 [刘悦, 马舒畅, 杨正伟, 等. 面向材料领域机器学习的数据质量治理. *硅酸盐学报*, 2023, 51(2): 427–437]
- [105]. Du X, Lueer L, Heumueller T, et al. Revealing processing stability landscape of organic solar cells with automated research platforms and machine learning. *Infomat*, 2024, 6(7): e12554

- [106]. Sheng Y, Wu Y, Yang J, et al. Active learning for the power factor prediction in diamond-like thermoelectric materials. *npj Comput Mater*, 2020, 6(1): 171
- [107]. Lo S, Baird S G, Schrier J, et al. Review of low-cost self-driving laboratories in chemistry and materials science: the “frugal twin” concept. *Digital Discovery*, 2024, 3(5): 842–868

AI-Driven New Materials Development and Data Standardization

Lanting Zhang^{1,2}, Mingyang Liu¹, Hongjian Yuan^{1,2,3} & Hong Wang^{1,2,3,4*}

¹ School of Materials Science and Engineering, Shanghai Jiao Tong University, Shanghai 200240, China

² Materials Genome Initiative Center, Shanghai Jiao Tong University, Shanghai 200240, China

³ Zhangjiang Institute for Advanced Study, Shanghai Jiao Tong University, Shanghai 201203, China

⁴ Suzhou Laboratory, Suzhou 251123, China

* Corresponding author, E-mail: hongwang2@sjtu.edu.cn

In recent years, with the rapid advancements in artificial intelligence (AI) and automatic technologies, AI-driven self-driving or autonomous laboratories have shown great potential in the research and development of new materials and are poised to become the "new infrastructure" for the paradigm shift in materials innovation. In this paper, the latest advancements in AI-driven self-driving laboratories (SDL) for accelerating new materials discovery are overviewed, focusing on the current status world-wide and the core challenges faced. Unlike automatic laboratories, a self-driving or autonomous laboratory features the design module, the preparation and processing module, the characterization module and the analysis and optimization module, which integrates laboratory automation, robotic technologies, AI algorithms and databases into a whole system, forming a closed-loop feedback workflow that efficiently optimizes target performance without human intervention. Depending on the hardware and software employed, the level of autonomy of SDLs can be classified from level 0 to 5, in which level 5 is the highest. Leading examples of SDL systems include the Mobile Robotic Chemist developed at the University of Liverpool, Berkeley's A-Lab, Autonomous MAterials Search Engine by the University of Maryland and China's AI Chemist platform, which have demonstrated remarkable success in accelerating materials discovery. In addition, the concept of asynchronous cloud-based delocalized closed-loop discovery to drive multiple geographically distributed platform has been successfully applied in the search for small-molecule gain materials for organic solid-state lasers.

Data-driven is the foundation of AI technology, in which both data quantity and quality are important. Data standardization emerges as a critical challenge and opportunity in SDL development. In addition to the level of autonomy, SDL holds the character of an intelligent data factory, which are not only the data producers but also the data users. The data standardization is the common language understood by the laboratories and is the key to breaking down information silos in self-driving

laboratories and transitioning from "isolated intelligence" to "collaborative intelligence". International efforts like OPTIMADE's API standard and NOMAD's metadata framework are paving the way for global data interoperability, with China contributing through its Materials Genome Engineering Data standards. The systematic development of a standardized materials data framework is currently underway in China.

Despite remarkable progress, SDLs still face significant challenges that must be addressed to realize their full potential. Current systems remain constrained by technical limitations, particularly their predominant focus on solution and powder chemistry while struggling to incorporate more complex processes, creating gaps in applications such as wide-bandgap semiconductors. While substantial automation has been achieved, critical dependencies on human expertise persist for cross-modal data integration and complex optimization tasks, preventing fully autonomous operation. The exorbitant costs of customized SDL platforms, often exceeding \$1 million, further restrict widespread adoption.

Looking ahead, the future development of SDLs will require multi-faceted advancements including expanded hardware capabilities to encompass vacuum-based techniques and bulk material processing, integration of more sophisticated AI approaches like reinforcement learning and generative models for true inverse design, implementation of cost-reduction strategies through modular "frugal twin" systems with open-source components, and establishment of comprehensive data standards following FAIR principles to enable seamless global collaboration.

This transformative approach has already demonstrated significant impact, such as reducing discovery timelines from centuries to weeks and achieving 6-fold improvements in catalyst performance. With continued advancement in AI integration and data standardization, SDLs promise to become the foundational infrastructure for next-generation materials innovation.

Keywords: artificial intelligence, autonomous or self-driving lab, materials design, data standardization, data-driven