

科技评论

AI 能否成为化学知识的发现者？

江俊^{1,2}, 崔乘幸^{2,3}, 黄文光⁴

摘要 人工智能(artificial intelligence, AI)正推动科学研究范式由“加速器”向“发现者”转变。以中国科学技术大学“机器化学家”平台的实践为例,系统分析 AI 在化学知识发现中的潜力与挑战。AI 通过机器学习、知识图谱和自动化实验系统,能够在分子设计、谱学分析、催化筛选和材料研发中实现从数据到知识的跃迁。然而,要使 AI 真正成为化学知识的发现者,仍有“三重瓶颈”亟需突破:高质量数据匮乏、人类认知维度受限及实验验证效率低下。进一步探讨了化学大模型、多模态数据融合及工业级智能实验室如何通过数据驱动决策优化、跨学科研究加速与自动化实验流程重构,推动未来科研范式的系统性重塑。

关键词 人工智能;机器化学家;化学大模型;可解释性;知识发现

1 为什么 AI 可以成为化学知识的发现者？

化学作为一门传统的实验科学,长期以来依赖于“试错法”进行探索。然而,随着化学空间的日益复杂化,穷尽所有可能性变得愈发困难,这导致了研发效率的显著降低。

当前 AI 的快速发展为科学研究带来了深刻的范式变革。借助机器学习、知识图谱和智能自动化平台,科研人员正探索 AI 赋能

的科学研究(AI for science)新路径。DeepMind 公司开发的 Alpha Fold AI 模型,通过训练海量蛋白质结构数据集,预测氨基酸间的距离和角度,实现蛋白质折叠结构的快速、高精度模拟,成果获得 2024 年诺贝尔化学奖^[1]。

国家层面的政策驱动为这一变革提供了制度支撑:2023 年科技部启动“AI for Science”专项,聚焦化学等基础学科;2025 年国家自然科学基金持续加大对“AI+基础学科”交叉研究的支持,推动科研模式向智能化、系统化转型。在此背景下,化学研究正加速迈向“人机协同、AI 主导”的新范式。通过融合化学知识、实验数据与先进算法,形成以数据驱动、模型引导、反馈优化为特征的智能科研流程。

以 GPT-4 为代表的大型语言

模型(LLM)的崛起让全社会看到了 AI 技术的新高度。在化学领域,面向化学的大模型成为新的研发热点。例如,上海人工智能实验室研发的 ChemLLM(Chemical Large Language Model)模型展示了高效处理复杂化学问题的能力。它可以帮助快速检索和总结文献、回答化学问题、辅助规划实验步骤。

另外,在科学计算方向,大模型的理念也开始延伸到分子模拟与材料设计中。中国“深度建模”(DeepModeling)开源社区于 2024 年发起了 OpenLAM(Open Large Atomic Model)计划,旨在通过对大规模量子力学数据的预训练,获得对元素、原子和分子系统的通用表示和预测能力,最终目标是能自行提出科学假设并完成模拟和实验设计。本质上,ChemLLM 和

1. 中国科学技术大学精准智能化学全国重点实验室,化学与材料科学学院,合肥 230026
2. 河南省科学院智慧创制研究所,郑州 451162
3. 河南科技学院化学化工学院(计算化学研究所),新乡 453003
4. 科技导报社,北京 100081

收稿日期:2025-07-07;修回日期:2025-10-19

基金项目:国家自然科学基金青年科学基金项目(A类)(22025304)

作者简介:江俊,教授,研究方向为计算化学和人工智能化学,电子邮箱:jiangjl@ustc.edu.cn

引用格式:江俊,崔乘幸,黄文光. AI 能否成为化学知识的发现者? [J]. 科技导报, 2025, 43(21): 16-22; doi:10.3981/j.issn.1000-7857.2025.07.00034

OpenLAM 分别代表了知识智能和实验智能 2 个方向的大模型探索,一个偏重“文科”(知识推理对话),一个侧重于“理科”(数值计算模拟),二者相辅相成。

中国科学技术大学江俊团队通过构建具有“感知—决策—执行—学习”闭环能力的“机器化学家”平台“小来”与“小临”,聚焦在分子合成、催化筛选、材料设计等领域的研发。例如,他们将火星陨石催化剂的筛选,从原本需 2000 年完成的工作被压缩至 5 周^[2]。这一效率跃迁背后,AI 正从数据处理工具向“知识合作者”进化,但它能否突破人类认知边界,成为独立的化学知识发现者?这一命题不仅关乎科研范式革新,更触及科学发现的本质逻辑。

首先,AI 作为辅助工具,可以帮助研究人员识别变量之间的潜在关联性,并进一步挖掘其背后的知识结构,这是当前 AI 在科学领域最擅长的方向之一。在这个层面,AI 通过对大规模实验或计算数据的学习,捕捉变量之间的统计或模式关联,并构建相应的关系网络或知识图谱,为后续的理论分析与机制探索提供切入点。Cui 等^[3]在 2024 年发表的论文中,以 CO₂ 红外/拉曼光谱为表征指标,定量评估电场强度对电催化性能的影响。以金属掺杂 g-C₃N₄ 单原子催化剂为例,理论探讨 27 种催化剂在不同电场下的吸附模式及能量变化。运用机器学习构建光谱描述符与吸附能量/电荷转移的智能模型,实现电场效应对 CO₂ 转化的定量化分析。注意力机制揭示光谱-吸附关联,并基于光谱逆向

预测电场强度,为机器学习光谱学监测调控电催化反应开辟了定量路径。AI 虽然可以为人类科学家指明研究重点与潜在规律,但是后续仍需依赖科研人员基于领域知识进行物理图像或数学理论的构建与验证。AI 正从工具转向“知识发现者”,未来化学研究将更加依赖于数据基础设施、智能模型与自动化平台的协同发展。

2 AI 能否胜任知识发现者?

人工智能具备综合性思考能力,能够跳脱传统认知边界,带来一些“奇思妙想”。然而,迄今为止,AI 在化学领域尚未产生真正颠覆性的原创理论,其背后主要有 3 方面的原因。

2.1 高质量数据匮乏

在当前人工智能赋能化学研究的探索中,数据仍是构建化学大模型的核心基石。然而,优质化学数据的获取与利用面临多重挑战。

高质量的数据匮乏是 AI 发展的主要瓶颈之一,尤其缺乏失败实验的数据,导致训练数据不完整并带有明显偏差。真正的高质量化学数据应具备功能标签,即通过实验手段对样本进行系统性的标注,明确其性质、功能或行为,从而提高模型训练的精准度与解释力。在化学大模型的构建中,常见的标签类型包括化合物实体识别、反应角色标注、分子性质标签、反应类型识别、图谱数据提取、材料合成路径等。这些标签大多仍依赖人工完成,而前提则是拥有原始的、结构清晰的数据。

现阶段 AI 尚未真正达到“集

成人类智慧”的高度,其训练数据主要来源于已发表的成功案例,尤其集中在反应合成路径上。化学反应理论上可能性极多,只要符合化学计量关系就可成立,但真正的挑战在于代价是否可接受、驱动是否可行。许多化学品的发现背后往往包含上万次试错,而这些失败的记录并未被系统性收集和学习^[4]。当前的 AI,仅整合了“片面的成功经验”,真正的智慧则需要失败中不断迭代和进化。

此外,现有化学相关数据库尚未实现全面开放。以全球最全面的商业化学反应数据库 Reaxys 为例,因授权限制,用户只能远程检索,无法获得本地化数据用于模型训练。这类数据库还普遍存在结构化不足的问题,其数据多为已发表的成功实验,缺乏失败记录及详细的功能属性描述,难以满足模型深层理解与泛化的需求。

化学数据的多模态融合也面临挑战:文本(文献)、图谱(分子结构)、光谱(NMR/IR)等数据异构性强。上海人工智能实验室开发的谱学大模型 ChemLLM 与深度势能的 DPA-2 模型虽具备潜力,但因缺乏高质量实验数据验证,实际效果受限。

2.2 人类认知维度受限

AI 被“投喂”高质量数据后,会产生一些有“见地”的认知。但并不被当下的研究人员完全理解,或者说 AI 在将它发现的知识向人类“传授”时,对知识进行通俗化的表达(即可解释的理论),不可避免会丢失了本来的“味道”。

信息时代,研究人员亟需对可解释性的本质进行反思。当前人

类所谓的“可解释性”，在某种程度上是一种迎合人类认知习惯的“降维表达”。人类对科学规律的理解，主要依赖于2种形式：一是数学公式，即可由变量间解析表达式表示的关系；二是可视化图像，即将变量间趋势用二维或三维图形展现。但现实中，AI所学习到的规律往往位于高维特征空间，可能涉及七八个，甚至十几个变量之间复杂的非线性关系。为了使这些高维规律为人类所理解，AI必须进行“降维处理”——要么将其转化为较为简单的数学表达式(如主成分方程、回归模型等)，要么可视化为图表趋势。但这种降维本身就会牺牲信息、扭曲原貌，仅保留可理解部分，并不能还原AI原本所掌握的全部知识结构^[5]。高维空间中变量的协同作用，可能根本无法在二维图像或单一表达式中直观展现，因此，这种“可解释性”在本质上是一种妥协式的理解方式。

目前，在许多尚未建立清晰理论体系的前沿领域，AI所揭示的复杂关联关系甚至超出了人类当前的理解与验证能力。这并不是AI的“神秘”，而是反映了尚未形成足够成熟的分析工具。因此，回顾历史可以发现，人类每一次科学范式的跃迁都伴随着方法学工具的革新。要想更好地与AI协同，人类首先必须发展并掌握更先进的认知工具。例如，麦克斯韦方程组虽然在形式上未必像某些数学定律那样“优美”，但它构建了一个极其完备的电磁理论框架，是理解电磁现象的基础。这种形式化的数学工具，代表了人类对自然规律的深度抽象能力。因此，推动

AI与科学研究的融合，首要层次是人类要具备解析AI输出的能力，并通过数学建模等手段将其转化为可验证的科学表达。

2.3 实验验证效率低下

目前，AI提出的假设和预测大多需要通过传统实验手段进行验证，这种方法低效、耗时，一定程度上阻碍了AI科学假设的快速迭代和验证。

以燃料电池催化剂的配方优化为例，AI提出了组合方案，按照该方案进行的实验起初表现平平，若按传统思路容易被放弃。然而，研究团队持续优化后发现：只有在控制3种添加组分中锌含量极低的情况下，催化性能才显著提升。也就是说，AI的思路在传统试验模型中可能并不总能得到充分展现。现有的一些化学大模型虽能处理复杂谱学数据，但因缺乏实时反馈的自动化实验平台，其预测难以快速迭代，形成“设计—验证”的效率断层。因此，当下急需发展具有快速响应和智能执行能力的“行动智能”系统，与AI的设计能力相匹配。

3 AI知识发现的可靠性、数据基础与人机协同

3.1 AI发现的化学知识是否可靠？

深度模型往往基于大数据中隐含的统计相关性进行学习，而非建立在真实的因果推理基础上。这种“表面相关”的逻辑一旦脱离原始数据分布，很容易导致预测失真，产生不合理，甚至物理上违背常识的结果，例如，生成违反热力学定律的反应路径。因此，如何提

高模型可解释性，清晰揭示预测背后的科学逻辑，是AI真正成为知识发现者的关键。

Wang等^[6]系统阐述了“以可解释性AI为桥梁”破解上述困境的有效路径。该研究聚焦于金属-载体相互作用这一催化领域的核心科学问题。此类相互作用对催化剂的活性、选择性与稳定性具有决定性影响，但由于其依赖因素复杂(包括金属与载体的组成、尺寸、形貌及反应条件等)，长期以来其本质机制一直处于“黑盒”之中。为突破这一瓶颈，该研究团队^[6]首先系统整合了文献中已有的界面作用实验数据，涵盖了25种金属和27种氧化物。随后，研究人员基于基础材料属性构建了一个包含约300亿个潜在组合表达式的超大特征空间。通过引入具备良好可解释性的压缩感知算法，结合化学领域知识与理论推导，最终筛选出具有明确物理意义、计算结果准确的关键描述符，并建立了控制金属-载体相互作用的数学方程。该方程首次揭示了决定该作用强度的两个核心因素，并具有良好的普适性与可推广性。

基于可解释性算法的AI模型，能够从复杂实验数据中抽象出物理规律，建立具有预测能力的数学理论，为揭示科学机制、发现潜在法则提供了可行路径。这种方法不仅提升了模型的可信度与可控性，也有望推动AI从“辅助工具”向“知识发现者”角色的根本跃迁。然而，在构建类似数学模型或公式组时，其结构与复杂度往往超出人类直观理解的范围，仍需进一步发展人机协同的研究范式，让

科学家在保持对原理理解的同时,有效借助 AI 挖掘潜在规律。

实践是检验真理的唯一标准。AI 发现的化学知识,最终仍需要在实验中验证其可靠性。未来,随着自动化实验平台和智能实验室的普及,AI 生成假设—自动化实验验证—模型迭代优化将形成闭环,显著缩短从“发现规律”到“验证真理”的周期。

3.2 如何获取高质量的数据提升 AI 的智力?

针对化学数据稀缺、标注昂贵的问题,中国科学技术大学研究团队^[7]提出了“理论大数据+实验小数据”的融合策略:一方面,通过高通量理论计算模拟生成大量数据用于模型预训练;另一方面,再用少量真实的实验数据进行精调,有效提高模型的泛化能力与实际可用性。要推动 AI 真正引领化学研究范式变革,必须以高质量、多模态、结构清晰的数据体系为基础,配合先进的模型训练策略与开放的科研生态,构建更加智能、精准、可解释的化学大模型。

在数据的可获取性方面,中国具备相较于其他国家更为突出的优势。受限于隐私保护与知识产权法规,国外较难获得具备全属性的数据。相比之下,中国的数据政策更具开放性。通过完善制度设计与激励体系,可鼓励科研人员利用电子实验记录本(ELN)等工具,自主上传实验数据,特别是未发表的失败案例与边界探索成果,为化学大模型提供更广泛的学习素材。

此外,化学大模型通常面临多模态数据的融合挑战。化学数据类型涵盖文本(如文献与专利)、

结构图(分子式与反应式)、图像(如晶体结构图、显微照片)、光谱(如 NMR、IR、XPS)、序列(如聚合物链或反应路径)等多种形式。这些数据异构性强、格式复杂,如何统一编码与融合,构建可被 AI 理解与利用的知识图谱,是当前模型构建中的一大难题。

近期的技术进展为多模态融合提供了新的路径。2025 年 4 月 16 日,上海人工智能实验室发布并开源了通用多模态大模型“书生·万象 3.0(InternVL3)”。该模型采用创新的原生多模态预训练方法,在预训练阶段即将文本与视觉信息无缝结合,区别于传统“语言模型+视觉插件”的后融合方式,使模型具备原生处理多模态输入的能力。在多个专家级评测中,其 10 亿~780 亿参数的全系列版本在开源模型中均表现优异。

3.3 人类如何适配 AI 快速发展的脚步?

随着科学体系的持续膨胀,人类个体已难以掌握所有知识的细节。许多知识只能被“封装”在模型、算法或技术系统中,人们无须理解其全部内部机制,只需要理解其输入与输出之间的基本逻辑。这种“功能性理解”在现代社会已无处不在:每个人都能高效使用智能手机,但极少有人真正掌握其底层芯片设计、电信协议或操作系统内核的全部细节。

在 AI 时代,知识封装成为必然趋势。面对无穷无尽的知识总量,个体的时间与精力终究是有限的。正如古人所言:“吾生也有涯,而知也无涯。”因此,人类应借助 AI 的力量,在“有限生命”中聚

焦真正有意义的科学难题,把有限的认知资源用于探索未知、挑战边界。AI 与人类的协作应基于双重路径:一方面,人类必须通过数学与理论工具提升自身解析 AI 输出的能力,建立可验证的科学体系;另一方面,应合理接受知识封装机制,在必要时将 AI 模型作为可信赖的“黑箱工具”,从而将精力集中于更具创新性和突破性的科研任务。

真正具有科学创造能力的智能体,需具备精准理解并预测物质世界的能力。而衡量其水平的关键,仍是实验室中“能否验证”。短期内,实验仍将是不可替代的验证手段。但随着智能体理解力与硬件能力的同步提升,人类有望迎来从“实验验证”向“智能预测”的科学跃迁。

4 AI 重塑科研范式:基础设施、应用赛道与探索边界

4.1 工业级智能实验室将怎样重塑科研范式?

与“机器化学家”这一智能科研平台相呼应的,是实验空间形态的深刻变革。长期以来,化学实验室多呈现出“手工作坊式”特征,设备配置不一、操作流程非标准化、数据质量难以统一。这种模式在现代科研高密度、高通量的需求下,已显现出明显的不适配。

当前科研效率的瓶颈之一,正是实验系统的非工业化。世界正处于高度自动化的大工业时代,而科研实验室却仍停留在依赖人工经验的传统模式。未来的实验室将迈向工业级智能体集群协作体

系,实现设备、流程、数据的全面标准化与智能化。

在这一愿景下,自动化合成单元、高通量表征平台、在线数据采集与分析系统将通过统一的操作系统与调度平台联接成“实验工厂”。有机合成、电化学、高分子材料等方向可以建设各自的专业子平台,实现从反应设计、条件筛选到产物分析、性能评估的全流程自动化。智能机器人逐步具备“感知型”能力,能通过光谱、视觉、温度与压力信号“看见”和“感受”反应过程,从而在微观层面理解实验进程。

在更大尺度上,多平台、多实验室通过云端互联形成“科研智能体网络”。科学家只需在远程终端上发出指令,即可调用云端机器人完成高复杂度实验任务,实现真正意义上的“离身科研”和“远程创新”。研究人员从大量繁琐的重复性操作中解放出来,将更多时间投入到问题抽象、机理思考和结果解读之中,科研活动从“体力密集型”向“智力驱动型”彻底转变。

4.2 AI助力下,化学首先会在哪些赛道实现突破?

人工智能正以前所未有的深度介入化学的多个二级学科,其应用已从早期的辅助工具转向推动学科变革的核心动力。目前,AI在化学领域的典型应用包括分子设计、反应预测、催化剂筛选和合成路径规划等方向,并在多个研究环节取得了一系列具有突破性的成果。

从学科本身的发展逻辑出发,理论化学可能是最先全面受益于AI加持的领域。理论化学的核心瓶颈在于其计算结果与实验观测

之间存在偏差,导致可验证性差、可信度不足。AI的介入为这一难题提供了2方面的突破。

首先,AI显著加速了理论计算的过程,能够处理大量复杂体系的模拟与推演,从而扩展了理论的可探索边界;AI作为接口,将理论数据转化为具备实际意义的网络模型,使得理论结果可与实验数据直接对接,形成“可验证—可迭代—可优化”的闭环机制。在这一框架下,科研人员可先训练出一个具有一定准确性的预训练模型,用于探索潜在的物性规律。再用零散的实验数据对模型进行动态修正和快速演化。

其次,配方筛选类问题也是AI较容易突破的方向,尤其适用于具备高通量筛选需求的化学学科。此类问题通常具有不超过10个输入参数和3个左右输出目标的有限维度,正好适配当前AI模型在样本复杂度与模型拟合能力之间的平衡点。催化化学和合成化学是2个典型受益方向——前者可用于高效筛选催化剂组合,后者则可借助AI在反应路径规划中寻找生成目标产物的最优路线。例如,华东师范大学姜雪峰教授团队通过定制化逆合成AI策略,建立专用成环模板库并筛选高多样性路线,实现对多取代对苯二酚甲醛(PGAs)的一步构建。AI辅助设计结合自动化流体合成,不仅优化了成环选择性和反应条件,还显著提升了产率与反应效率。通过DFT(密度泛函理论)计算阐明炔基锂试剂的化学选择性及1,5-HAT机制,AI在多样化策略探索、路线筛选和快速反应优化中

发挥核心作用,为PGAs(聚甘油脂肪酸酯)工业化制备提供高效可扩展方案^[8]。

在实际应用与商业化方面,药物合成化学是当前AI最具产业潜力的方向之一,尤其是与蛋白质功能相关的研究领域,已成为AI制药公司与技术平台争相布局的热点。通过AI辅助分子设计、靶点匹配与合成优化,药物研发周期正逐步从数年压缩至数月,带来了切实可观的经济价值与社会效益。

总体来看,与理论建模联系紧密、结构变量有限的化学子领域,在当前AI发展的阶段更容易取得实际突破。同时,AI与化学的发展也呈现出相互促进的态势。AI为化学研究提供了强大的数据建模与决策优化能力,化学则为AI提供了丰富且具体的应用场景与产业化通道。

4.3 科学探索的“偶然之美”还能出现吗?

“机器化学家”系统“精准理性”的背后也引发了新的思考:历史上许多具有突破性意义的科学发现,恰恰源自一次偶然的实验失误或非典型现象。例如,青霉素的发现、石墨烯的剥离,甚至量子力学早期的某些观察,皆带有偶然性和不可预设的成分^[9-10]。那么,在强调可靠性与再现性的智能科研体系中,又该如何保留科学探索的“偶然之美”?

对此,当前的学术研究普遍强调要“顶天立地”——既要追求理论上的突破与前沿,也要解决现实中的具体问题。然而现实中,这两者往往难以兼顾。真正革命性的发现虽然令人振奋,但从资源投入

与科研生态的角度看,并非每一位科研人员都能持续进行高风险、高回报的研究。科学研究需要定位清晰。如果发现了自己的长处,就应该在这个方向上深耕、主动发力。

在这一背景下,人工智能的现阶段能力,应该更多聚焦于“立地”的方向,即将既有理论和知识体系转化为可落地、可应用的科研成果。100年前,爱因斯坦、玻尔、薛定谔等伟大科学家已经建立了人类物理和化学的理论基础。今天,更需要把这些理论变成真正能用的东西,这才是科研智能化最迫切的目标。

以人工智能在化学中的应用为例,当前许多成功的案例正是“先解决问题,再探讨机理”的典型路径。例如,AI在数据中识别出某些变量之间的相关性,尽管这一过程并不以已知科学原理为基础,但这种相关性在帮助科研人员有效解决了实际问题后,便可进一步反向解析其背后的科学规律,最终导向理论上的新突破。这其实是一种新的科研逻辑路径:由数据驱动的问题解决出发,逐步过渡到原理揭示与机制创新。在智能实验体系下,探索性并不会被精准性完全排除。通过合理引入“扰动机制”或“目标偏离策略”,AI系统也可以在实验中引入受控的随机性,形成类似人类科研中的“灵感瞬间”。

未来,AI不仅是精密执行者,更可能成为主动探索者。但前提仍然是在大量高质量的“立地”型研究中积累数据与经验,夯实基础认知体系。只有当AI拥有足够“理解世界”的能力时,科学研究

才可能从“实验验证”迈向“理论自洽与直接预测”的全新范式。

5 AI时代的化学人才与科研角色重构

5.1 AI时代,化学创新人才的成长路径何在?

在传统语境中,“传道、授业、解惑”是教育的基本使命。但在人工智能深度介入教育与科研的当下,这3者的权重正在悄然发生变化。在授业和解惑方面,AI的能力在许多情境下已展现出明显优势。AI背后连接的是全人类的知识体系,几乎可以做到即时反馈、精准讲解。在这种情况下,教师的核心价值应当转向“传道”,即以科学的方法论与思维方式,引导学生形成正确的科研理念和价值判断。

AI或许能发现上千个可行的公式与路径,但真正具备判断力的,仍是人类自身。在这些可能性中,哪个方案最具科学意义,最能解决关键问题?这是需要经验、直觉和科学品位来甄别的。这种品位的形成,不仅是知识的积累,更是方法论的传承。未来导师的角色,将更多体现在对学生科研品位的塑造与学术判断力的培养上。

中国科学技术大学李群祥教授谈到讲授量子物理时经常强调,“前几课不要急于讲公式,而要先讲科学史。讲黑体辐射是如何启发普朗克思考能量量子的,氢原子谱线又如何引发玻尔对原子模型的构想。”这些经典案例并不是单纯传授知识点,而是引导学生理

解科学家如何面对未知、如何提出问题、如何构建模型,这才是科学教育中最核心的“道”之所在。

在AI辅助科研日趋普遍的背景下,未来的研究人员将愈发“原子化”——每个人的研究风格、学术趣味和问题选择将变得更加多元和个性化。与围棋不同,科学研究的胜负并不只有一个标准,问题解决的路径更是千差万别。

5.2 AI进入科学领域,会否让研究生失业,甚至让科学家无事可做?

现阶段,AI的核心价值在于助力科研人员高效应对高维参数的筛选与优化。然而,物质科学的疆域广袤无垠,真正的突破始终离不开人的思想、灵感与直觉。“机器化学家”平台的系统设计中,刻意保留人机协同(human in the loop)的关键环节,让化学家在问题定义、结果判读和研究方向选择等核心节点深度参与决策。

因此,AI进入科学领域所催生的“智能科学家”,其使命从来不是取代人类,而是赋能并增强科学家的能力。担心AI与机器人会让研究生失业,就如同当年经典物理学家以为“科学大厦即将封顶”,严重低估了物质科学的无尽前沿。历史反复证明,每一种新工具的诞生,目的都是提升人类的探索力,结果也必然是带领我们开拓更广阔、更未知的天地。AI不是终点,而是通向未来的加速器。对于化学领域的青年研究者而言,更重要的不是“是否会被AI取代”,而是“如何与AI共同成长”,让自己成为那个能驾驭智能工具、引导科研方向的人。

6 结语

人工智能正在深刻重塑化学研究范式。“AI+化学”不仅显著提升了科研效率,使原本耗时千年的探索在数周内完成,而且打通了理论与实验的壁垒,为科学家深入理解复杂体系提供了新路径。在新材料、新能源、生物医药等前沿领域, AI 正成为驱动原创性突破的关键引擎。

然而,这一变革仍处于早期阶段。其进一步发展面临诸多挑战:高质量、标准化数据体系尚未健全;模型的可解释性与泛化能力仍有限;多模态、多尺度系统的协同建模尚不成熟。同时,科研管理与教育体系也需适应变化,推动政策机制更新,例如,促进数据共享、跨学科合作,改革科研评价体系,培养兼具化学背景与 AI 素养的复合型人才,并建立 AI 实验室的伦

理与安全规范。

AI 不应被神化或简化,而应与人类智能协同互补。AI 擅长数据处理与模式识别,人类则在问题定义与理论抽象中具备不可替代的创造力。二者融合,将成为推动科学跃迁的重要力量。

参考文献(References)

- [1] Jumper J, Evans R, Pritzel A, et al. Highly accurate protein structure prediction with AlphaFold[J]. *Nature*, 2021, 596(7873): 583–589.
- [2] Zhu Q, Huang Y, Zhou D L, et al. Automated synthesis of oxygen-producing catalysts from Martian Meteorites by a robotic AI chemist[J]. *Nature Synthesis*, 2024, 3: 319–328.
- [3] Cui C X, Shen Y X, He J R, et al. Quantitative insight into the electric field effect on CO₂ electrocatalysis via machine learning spectroscopy[J]. *Journal of the American Chemical Society*, 2024, 146(50): 34551–34559.
- [4] Maloney M P, Coley C W, Genheden S, et al. Negative data in data sets for machine learning training[J]. *Organic Letters*, 2023, 25(17): 2945–2947.
- [5] Barredo Arrieta A, Díaz-Rodríguez N, Del Ser J, et al. Explainable artificial intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI[J]. *Information Fusion*, 2020, 58: 82–115.
- [6] Wang T R, Hu J Y, Ouyang R H, et al. Nature of metal-support interaction for metal catalysts on oxide supports[J]. *Science*, 2024, 386(6724): 915–920.
- [7] Xiao J, Zhang Y F, Li B W, et al. A deep learning-augmented density functional framework for reaction modeling with chemical accuracy[J]. *JACS Au*, 2025, 5(8): 3892–3903.
- [8] Meng J L, Yang H B, Li C L, et al. Machine learning-driven automated synthesis of polysubstituted gentaldehydes[J]. *Angewandte Chemie International Edition*, 2025: e202515595.
- [9] Fleming A. On the antibacterial action of cultures of a *Penicillium*, with special reference to their use in the isolation of *B. influenzae*[J]. *Clinical Infectious Diseases*, 1980, 2(1): 129–139.
- [10] Novoselov K S, Geim A K, Morozov S V, et al. Electric field effect in atomically thin carbon films[J]. *Science*, 2004, 306(5696): 666–669.

Can AI become a discoverer of chemical knowledge?

JIANG Jun^{1,2}, CUI Chengxing^{2,3}, HUANG Wenguang⁴

1. Key Laboratory of Precision and Intelligent Chemistry, School of Chemistry and Materials Science, University of Science and Technology of China, Hefei 230026, China
2. Institute of Intelligent Innovation, Henan Academy of Sciences, Zhengzhou 451162, China
3. School of Chemistry and Chemical Engineering (Institute of Computational Chemistry), Henan Institute of Science and Technology, Xinxiang 453003, China
4. Editorial Department of Science & Technology Review, Beijing 100081, China

Abstract Artificial intelligence (AI) is driving a paradigm shift in scientific research—from functioning primarily as an “accelerator” to emerging as a genuine “discoverer.” Using the “Machine Chemist” platform at the University of Science and Technology of China as an illustrative example, this work provides a systematic analysis of the potential and challenges of AI in chemical knowledge discovery. Through machine learning, knowledge graphs, and automated experimental systems, AI can achieve a true transition from data to knowledge in areas such as molecular design, spectroscopic analysis, catalyst screening, and materials development. However, for AI to become an autonomous discoverer of chemical knowledge, three critical bottlenecks must be addressed: The scarcity of high-quality data, the limitations of human cognitive frameworks, and the low efficiency of experimental validation. This study further examines how chemical foundation models, multimodal data integration, and industrial-scale intelligent laboratories can drive systematic transformation of future scientific research paradigms by enabling data-driven decision optimization, accelerating interdisciplinary research, and restructuring automated experimental workflows.

Keywords artificial intelligence; machine chemist; chemical foundation model; interpretability; knowledge discovery ●



(责任编辑 王微)