DOI: 10.7544/issn1000-1239.qy20260701 CSTR: 32373.14.issn1000-1239.qy20260701

前　言

——AI4Science: 从科学助手到自主科学家

对科学研究来说，重要步骤之一是“提猜想（hypothesis）”——爱因斯坦曾说:“仅看提出问题并清晰阐述之，即已远重于问题之求解（The mere formulation of a problem is far more essential than its solution）”；李凯、李国杰等计算机学者亦有此论。

从这个角度而言，科学研究的确正经历着科研范式（research paradigm）的转变。过去的科研范式，诸如古希腊泰勒斯践行的观察与思辨、欧几里得集其大成的逻辑推演、伽利略与培根倡导的观察与实验、开普勒肇始的大数据分析与计算，其实施主体都是“人”，即：“人”来提猜想，“人”来开展实验，“人”来证实或证伪猜想；然而在 AI4Science 范式中，是“人机协同，甚至是机器自主地提猜想、开展实验”。科研活动的实施主体发生了变迁，称之为“科研范式的嬗变”非为过当。

我们先来看使用人工智能（artificial intelligence, AI）作为“科学助手”辅助人类科学家求解问题这一层面。AI 技术，本文特指当前常用的深度神经网络、大语言模型和智能体技术，之所以能够辅助人类科学家求解问题，部分源自如下几方面的能力。

1）大数据处理能力。现代科学（高能物理、天文学、基因组学、气候科学和材料科学等）实验装置产生数据的一个特点是“规模大”。例如，CERN 的大型强子对撞机每秒产生约 1 PB 原始数据；我国的“天眼”FAST 射电望远镜每年产生 15~20 PB 数据；全球基因组数据总量已达到 EB 级别。从如此大规模的数据中寻找规律，早已超出人类直接分析的边界，而这恰恰是 AI 技术的强项。大数据处理能力非为 AI 技术所独有，统计机器学习技术也有此能力，相关示例亦有很多，故在此不赘。

2）表征学习能力。现代科学装置产生数据的另一个特点是“维度高”，且各观测维度之间关系复杂。Leo Breiman 在《统计建模：两种文化》一文中所说：如果采用传统的统计建模分析数据的话，需要“事先假设数据的产生模型”；这种模型往往注重数学形式上的优美以便于数学推导，然而和数据的真实产生机理并不一致。因此，Leo Breiman 建议采用“算法模型（algorithmic model）”，即将“数据产生机制视作未知”，而采用机器学习技术直接从数据中学出“隐含模式”（或者说潜在规律）。粗略地说，这些机器学习技术是在做表征学习，即“把高维、复杂数据映射到低维流形，压缩到低维隐空间”，进而在隐空间中发现模式。

在此方向的范例之一是 DeepMind 团队与牛津大学的数学家 Marc Lackenby、András Juhász 教授合作的一项工作：先产生大量数据，然后使用监督学习及归因技术，发现了扭结的代数不变量与几何不变量之间的关联模式。这是以计算技术辅助人类科学家形成数学猜想的一次成功尝试。

3）跨学科综合和联想能力。现代科学的分支越来越多，越来越细。一位科研人员在一个学科有深厚积累已属不易，更遑论精通多个学科——如同冯·诺依曼、理查德·费曼那样的通才型科学家（polymath）益发罕有。

这种学科之间的壁垒以及博学的困难，导致“联想（association）”这一探索未知的重要手段有失效之虞。伟大的匈牙利数学家 G. Polya 将“联想”归入“合情推理”（plausible reasoning）的范畴，极为推崇“联想”的作用：如果发现了两个不同现象或不同领域之间存在相似性，即可将已知领域的成果推广到未知领域，从而大大推进对未知领域的探索。

与人类科学家不同，大语言模型以近乎所有的、文本形式的人类知识为训练语料，将不同学科的文本映射到同一个隐空间，故而可能发现“有些概念，即使源自不同学科，其本质是相近的”，可能发现那些因学科壁垒而长期被遮蔽的关联。

一个典型示例是询问 DeepSeek“二叉树叶子节点计数”问题，DeepSeek 不仅能够用计算机科学里的递归过程来解答，还能够联想到概率论中的随机游走的首次通过问题、分支过程的灭绝概率、双曲动力系统中的逃逸时间。这种跨学科的综合和联想能力，人的学养难以达到，恰恰是大语言模型的一个显著优势。

4）“智能”搜索能力。在可行解空间、猜想空间进行快速搜索，又是 AI 技术远胜于人的一个层面。典型示例之一是 AlphaGo：采用蒙特卡洛树搜索技术发现有利的棋局；为了缩小搜索范围，采用神经网络技术进行剪枝，指导搜索。后续的 AlphaDev 采用强化学习技术发现更高效率的排序算法，AlphaTensor 发现高效的矩阵乘法，FunSearch 利用大语言模型逐步改进并发现高效的算法，其本质皆可视为 AI 引导的高效搜索。

值得指出的是：搜索是计算机的强项，远胜于人是应有之义。上述这些工作的特点不仅仅是搜索，而是搜索时的“启发式规则”从哪里来。过去常常依赖于人的灵感、专家的经验、依据少数实例上的观察来设置启发式规则，自然难以称得上适用性强；上述工作则有所不同，采用强化学习、神经网络等技术，从大量数据或者反馈中自动学习出来，从而能够获得泛化能力强、适用性广的启发式规则。

在上述这些工作中，AI 是扮演“科学助手”的角色，仅仅完成完整科研过程中的某一个或几个环节，在单点上展示出超越人类科学家的能力，而对整体科研过程的设计与把控，依然掌握在人类科学家手中。

比上述工作更激进的是，AI 开始扮演“自主科学家”的角色，即：自主完成“提猜想—设计实验—完成实验—分析数据—修正猜想—获得发现”完整的科研过程。典型示例包括：

1）Sakana 团队设计了“AI 科学家”（AI scientist）系统，从人类科学家给定的两个种子思路（seed idea）出发，扩展出 10 个思路，并开展实验，分析结果，撰写文章。

2）DeepMind 团队设计的 Co-scientist 系统则聚焦于“提猜想”：以一个智能体根据用户输入的研究目标搜索文献，形成初步假设；以“反思”智能体评估、遴选假设；生成研究计划并执行。Co-scientist 系统在药物重定位等跨领域的复杂问题上初现威力。

3）FutureHouse 团队提出的 Robin 系统与 Co-scientist 类似，也是采用多智能体技术完成完整的科研过程，在干性年龄相关性黄斑变性研究中发现“增强视网膜上皮细胞的吞噬功能”是可能的突破口，进而发现已有的治疗青光眼的药物可能适用，并经实验证实药物的有效性。

值得指出的是，AI技术不仅可以推动“干实验”，还可以与机器人技术结合，形成自动化、智能化的实验设备，显著提高“湿实验”的效率与通量，减少实验员的人工劳动，催生“自驱动实验室”（self-driving laboratory）这一新模式。

这些进展让人惊讶，然而与此同时，我们也应该看到一些困惑、担忧甚至警示。

1）如何将AI的发现转化为人能够理解的知识？这是否是必要的？

正如Leo Breiman所说，“算法模型”（algorithmic model）抛弃了“刻画数据产生机制”这一要求和束缚，并因此而获得能力的显著提升。这自然也带来了一些不足之处，比如：如何将隐空间中发现的潜在模式翻译成人能够理解、验证的知识？

李明等将之称为AI技术的“投射问题”：潜空间中的模式要成为科学知识，需要被翻译成人类能理解和检验的概念；这个翻译过程，是一个需要创造力的过程。

典型示例之一是蛋白质结构预测算法AlphaFold。AlphaFold对许多蛋白质的结构预测精度达到了实验测定技术的水平；然而，AlphaFold仅仅是预测蛋白质结构，却并未揭示蛋白质的折叠机理，即：蛋白质为何会自发形成其天然态构象。

另一个示例是著名数学家M. Freedman等关于”人类数学（human math）与形式数学（formal math）对比”的研究。M. Freedman等分析了MathLib中的14万个数学定理，发现数学定理的“表述长度”通常很短，与数学定理“扩展后长度”呈现出典型的幂律关系。之所以会形成“短的表述”，M. Freedman等将其归功于“人类数学家不断总结、归纳出新的定义、术语和概念“，即：将隐空间中高频出现的模式翻译成人类数学家能够理解的知识和概念；而这种总结、归纳出新知识、新概念的能力，是当前的大语言模型所缺乏的。

但更为重要的是：这个问题本身是否成其为一个问题？换句话说，我们有无必要完全理解机器的智能？这依然尚无定论。

2）如何避免AI指导科研活动的趋同性？

完全依赖于AI技术指导科研过程，或者说一个完全自主运行的AI科学家，存在”模式塌缩“的内在隐患：AI指导下，科研的选题，或者说猜想的形成，有趋同化的倾向；科研的多样性有降低的倾向。

例证之一是2026年J. Evans等对4 130万篇科技论文的分析：他们发现，AI技术的使用将论文数量提升了3.02倍，引用数提升了4.84倍，然而科研选题的多样性则降低了4.63，科研合作降低了22%。AI技术倾向于建议那些“数据多、已有成果多的选题”，而不是有新意的选题；大语言模型通常输出那些“共识”性知识，但是较少输出那些长尾分布尾端的、非共识的，然而更具有新意的思路。换句话说，从科研群体来看，选题（或者说猜想）的多样性在降低，真正的新思维有被“慢慢挤走”的风险。

李明等认为，要想避免这种趋同性，“人机协同”是可能的解决方案之一：自主科学系统的效率来自把”人“的角色排除在外，而其风险也来自于此。一个真正有生命力的AI4Science系统，应当是人机协同的，即：AI执行提出猜想、搜索猜想空间、实验执行、数据分析等任务，而人类科学家始终参与问题设定、将隐空间的部分模式翻译成人类能懂的概念、价值判断等。如此保持人机协同过程的开放性，从而避免“开放的科学探索”退化为“封闭的目标函数优化”。

为深化对AI4Science利弊两方面的认识，我们特组织了本期专题，内容涵盖电路设计、计算机算法、农学、生物学、数学、材料科学等多个方面。刘睿之等所著《人工智能辅助的芯片物理设计方法综述》总结了如何使用AI技术提升芯片物理设计（比如布线问题）的效率；隋京言等所著《利用深度学习辅助算法设计——以排课问题与旅行商问题为例》介绍如何使用AI技术学习出“人类算法学家设计算法时的灵感”；王耀君等所著《气象大模型在农业应用中的综述》梳理了气象大模型的技术演进与分类体系，指出了当前模型在泛化能力、物理一致性等方面的挑战；王耀君等所著《基于Evo2模型微调的主粮作物基因组序列生成模型构建与应用》阐述如何应用大语言模型对马铃薯、花生等六种作物进行基因组序列生成；温翰等所著《人工智能在RNA研究中的应用》总结了人工智能在RNA研究中的核心应用，剖析了数据匮乏和模型可解释性等挑战，并从干湿实验闭环等角度提出了未来发展路径；贾伟乐等所著《机器学习原子间势的综述：模型、训练及评估》总结了机器学习原子间作用势函数（MLIPs）在模型架构、训练优化及模型评估三方面的最新进展，为衔接微观量子力学与宏观物性模拟提供系统性参考。这些文章不仅展示了AI技术辅助科学研究的成功，也展示了目前的困惑与思索。

当前AI技术发展迅猛，很多问题尚未有定论；在此种情景之下，唯有适时的总结、深入的思考与理性的争鸣，方为深化认识的可行之道。我们希望以此专题为引，激发读者诸君的思考、辩论与争鸣。

是为引。

卜东波（中国科学院计算技术研究所）

2026年7月